Essays on Financial Economics

by

Yan Liu

Business Administration
Duke University

Date: ______________________

Approved:

____
Campbell R. Harvey, Supervisor

____
Ravi Bansal

____
Ian Dew-Becker

____
Andrew Patton

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Business Administration
in the Graduate School of Duke University
2014

Abstract

Essays on Financial Economics

by

Yan Liu

Business Administration
Duke University

Date: ______________________

Approved:

____
Campbell R. Harvey, Supervisor

____
Ravi Bansal

____
Ian Dew-Becker

____
Andrew Patton

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Business Administration in the Graduate School of Duke University
2014

Abstract

In this thesis, I develop two sets of methods to help understand two distinct but also related issues in financial economics.

First, representative agent models have been successfully applied to explain asset market phenomenons. They are often simple to work with and appeal to intuition by permitting a direct link between the agent's optimization behavior and asset market dynamics. However, their particular modeling choices sometimes yield undesirable or even counterintuitive consequences. Several diagnostic tools have been developed by the asset pricing literature to detect these unwanted consequences. I contribute to this literature by developing a new continuum of nonparametric asset pricing bounds to diagnose representative agent models. Chapter 1 lays down the theoretical framework and discusses its relevance to existing approaches. Empirically, it uses bounds implied by index option returns to study a well-known class of representative agent models — the rare disaster models. Chapter 2 builds on the insights of Chapter 1 to study dynamic models. It uses model implied conditional variables to sharpen asset pricing bounds, allowing a more powerful diagnosis of dynamic models.

While the first two chapters focus on the diagnosis of a particular model, Chapter 3 and 4 study the joint inference of a group of models or risk factors. Drawing on multiple hypothesis testing in the statistics literature, Chapter 3 shows that many of the risk factors documented by the academic literature are likely to be false. It also proposes a new statistical framework to study multiple hypothesis testing under test

correlation and hidden tests. Chapter 4 further studies the statistical properties of this framework through simulations.

Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

Symbols

$\kappa_i(\log M)$	The i-th cumulant of the log discount factor M
$\mathcal{N}(\mu, \sigma)$	Normal distribution with mean (vector) μ and variance (matrix) σ
$Possion(\omega)$	Possion distribution with intensity ω

Abbreviations

GEF	Generalized entropy function
MGF	Moment generating function
OTM	Out-of-the-money
ATM	At-the-money
FWER	Family-wise error rate
FDR	False discovery rate

Acknowledgements

I am grateful to all the people who have contributed to my graduate studies. Their help and encouragement has been invaluable to me.

To begin with, I would like to express my deepest thanks towards my advisor Campbell Harvey for his patience, advice and encouragement. I have deeply enjoyed our conversations. His insights and knowledge in finance has been the most valuable aspect of my time in graduate school. Cam's impact on my development goes beyond finance. His passion for research and his constant pursuit of perfection will have a permanent impact on my professional career and my personal life.

I thank my thesis committee members, Ravi Bansal, Ian Dew-Becker and Andrew Patton, for their time and support. They have provided invaluable advice and encouragement for the development of my thesis. I would like to especially thank Ravi Bansal, my first four years' advisor, for his support and mentoring throughout my studies at Duke. His devotion to research has deeply impressed me and sets the standard for my academic career. I also thank Tim Bollerslev for his advice and constant interest in me. I wish to also thank Hengjie Ai, Michael Brandt, John Graham and Lukas Schmid for their time, advice and help.

Graduate school is a bitter-sweet experience for most people. For me, it would be far more bitter without the help from friends. Jiaming Cen, Zhengzi Li, William McCartney, Jinghan Meng, Alexandru Rosoiu, Stuart Webb, Wei wei, Basil Williams and Heqing Zhu and many others have given me tremendous support. I would like to thank them for their time and help.

Finally and most importantly, I would like to thank my parents, Wencai Liu and Guoqin Zhang, for their love, encouragement and always being there when I need them the most.

1

Introduction

1.1 Index Option Returns and Generalized Entropy Bounds

Asset markets generate risk and return characteristics that continuously challenge our thinking. To rationalize market abnormalities, economists create models under a few generally accepted economic principles. These models are constantly scrutinized and possibly rejected with the advent of new empirical findings, and new models are again proposed to accommodate new findings. In this process, a few important diagnostic tools have been developed by the literature to restrict the behavior of a plausible model. For instance, under the basic no-arbitrage condition, Hansen and Jagannathan (HJ, 1991) construct bounds on the second moment of the stochastic discount factor for a given asset menu. This nonparametric bound provides a simple way to summarize asset market data and helps screen candidate discount factors. Snow (1991) extends their work by showing how to bound higher moments of the pricing kernel. Bansal and Lehmann (BL, 1997) and Alvarez and Jermann (AJ, 2005) derive restrictions on entropy, a separate metric on dispersion, based on the equity risk premium. Stutzer (1995) proposes an information bound that minimizes the Kullback-Leibler Information Criterion. These nonparametric bounds rely on the fundamental no-arbitrage condition and provide unique lens through which we can characterize asset market data, diagnose existing asset pricing models, and design new models to explain a larger set of empirical regularities.

I contribute to this literature by providing a unifying theory on non-parametric bounds. Starting from the no-arbitrage condition alone, I show the existence of a continuum of bounds that restrict the δ-th norm of the pricing kernel, with $\delta \in (-\infty, 0) \cup (0, 1)$. Next, I show that these bounds can be naturally interpreted as restrictions placed by an optimizing investor with a power utility function. In particular, I define an augmented return space with respect to a pricing kernel and show that an agent's portfolio choice problem based on this augmented return space imposes a constraint on moments of the pricing kernel. In a strict duality sense, I show that my approach is complementary to the Hansen and Jagannathan approach. Motivated by the similarities between my bounds and the BL/AJ entropy bound, I complete my bound spectrum by showing that the entropy bound is a limiting case of my bounds. Finally, I describe the nonparametric bound universe and discuss its exhaustiveness.

To facilitate the application of my bounds, I propose a new metric termed the *generalized entropy function*. It is a natural generalization of the entropy concept (Stutzer, 1996, Backus, Chernov and Zin, 2011, Hansen and Sargent, 2008) and encodes all the information of a pricing kernel. The system of new nonparametric bounds can then be brought in to restrict the generalized entropy function of the pricing kernel. Through Taylor series expansions, similar to Martin (2008) and Backus, Chernov and Martin (BCM, 2011), I show how various moments of the pricing kernel contribute to the generalized entropy function and more importantly, how weighted asset return moments provide information on the entropy function. An example featuring a finite state complete market economy is given to gain insights into the workings of bounds.

For empirical applications, I attack the pseudo problem in a well-known class of models. Rare disaster models, as pioneered by Rietz (1988) and recently rejuvenated by a sequence of papers by Robert Barro and his coauthors (Barro, 2006, Barro and Ursua, 2008 and Barro et al., 2009), use tail information to explain market abnormalities, in particular the equity risk premium. The inherent difficulty for this strand of literature is on how to measure events that only happen rarely. Similar to BCM, who gauge disaster models' performances against index option data along several metrics, I also use option data to infer tail information in the pricing kernel. Unlike BCM, I consider static portfolio strategies involving option returns and rely on nonparametric bounds to study the tail behavior. My approach is different from and advantageous over BCM in several respects. First, no specific assumption is made on the connection between macroeconomic fundamentals and the market portfolio. As a result, my results are robust to model misspecifications. Second, instead of fitting a parametric model and using it to summarize the option cross-section, I take the realized option returns as given and study their implications on the pricing kernel. This again alleviates the goodness-of-fit concern of empirical option pricing models. Finally, a formal statistical framework is developed to test model performances. This not only allows us to consider multiple assets simultaneously but also generates statistical significance for different model configurations.

Turning to the empirical findings, I first document the unique moment characteristics of trading strategies involving deep out-of-the-money (OTM) put options from a nonparametric bound perspective. Bounds implied by OTM puts universally dominate bounds implied by either the market index or risk-neutral straddles. This highlights the pricing of jump risks in OTM puts, and is precisely the type of information one needs to bear on models with tail risks. Next, I use data implied bounds to confront standard rare disaster models. As a first step, I mark up the permissible parameter region in which all nonparametric bounds are simultaneously satisfied. Through this process, the discriminatory power of the newly developed bounds stands out. More importantly, my bounds allow me to better distinguish alternative tail specifications with asset market data alone, offering a way to circumvent the pseudo problem for disaster models. Lastly, to take statistical uncertainty into account, I develop a formal testing framework that can accommodate multiple assets and different types of bounds. Under this framework, I reject the benchmark disaster model and a few alternative specifications.[1] Nonetheless, the model's ability of approaching asset market bounds does look impressive and I believe the idea of raising pricing kernel dispersion through tail distortions is promising. Taken as a whole, my results suggest more sophisticated specifications of disaster models, possibly through the time-dependency of disaster probabilities along the lines of Barro and Ursua (2008) and Watcher (2008).

1.1.1 Diagnosing Dynamic Asset Pricing Models with Generalized Entropy Bounds

Modern asset pricing theory makes tremendous progress towards reconciling asset market movements and macroeconomic fundamentals. Prominent examples, including the long-run risks, habit and rare disaster models, rely on predictable variations and/or higher order moments of macroeconomic quantities to explain market returns. As a result, the corresponding discount factors also feature time-varying state variables and/or skewed and heavy-tailed innovations. Despite their differences in macroeconomic dynamics and preference orderings, they share one common feature: the pricing kernel departs from log-normality unconditionally and displays higher order moments, especially over long horizons. For instance, the long-run risks model

[1] More specifically, I reject the benchmark disaster model in Barro (2006) and a few other parameterizations at 5% significance. As a comparison across different parameterizations, I find that models that feature larger and less frequent jumps are rejected more strongly than models that feature smaller but more frequent jumps.

highlights the fluctuating quantities of risks. When integrated over time, the unconditional pricing kernel becomes heavy-tailed due to the mixing of normal shocks.[2] The disaster model, on the other hand, relies on infrequent downside movements of fundamentals and thus displays skew by assumption. The question is: How do we measure higher order moments of the pricing kernel and, more importantly, how does financial market data bear on these measures?

I use the *generalized entropy* developed in Liu (2012) to capture higher order moments of the pricing kernel. As shown in Liu (2012), the generalized entropy takes a system of nonlinear power transformations of the pricing kernel to characterize its unconditional moments. Moreover, market returns provide robust restrictions on these moments through the *generalized entropy bounds*. To facilitate the study of dynamic asset pricing models, I extend Liu (2012) in two important aspects. First, I apply the generalized entropy to multi-horizon pricing kernels. Second, I use conditioning information to sharpen inference with the generalized entropy bounds. Both ingredients are necessary to differentiate candidate models by making full use of financial market data.

My research is related to a recent paper by Backus, Chernov and Zin (BCZ, 2011) who evaluate representative agent models using *entropy* and bond yields. Their entropy measure is shown to be a special case of the generalized entropy, and is hence termed the basic entropy. Taking both the long-run risks and habit models as examples, I argue that the generalized entropy is able to reveal important moment characteristics of the pricing kernel that is missed by the basic entropy. In addition, conditioning information allows me to directly test predictability assumptions, which are the key to leading asset pricing models.

Using the generalized entropy as a diagnostic tool, I examine dynamic asset pricing models. Without conditioning information, I find that the habit model by Chan

[2]See Hansen and Scheinkman (2013).

and Kogan (2002) cannot generate enough higher order moments (i.e., third moments and beyond) to meet the generalized entropy bounds with certain powers. The rejection is statistically significant. This finding agrees with BCZ but strengthens their results as their rejection is only true in mean. With conditioning information, I find mild statistical evidence to reject the Campbell and Cochrane (1999) habit model. In particular, dynamic strategies that exploit time-varying economic uncertainty imply utility gains that are too high for the Campbell-Cochrane model to reconcile. The long-run risks model survives both exercises.

1.2 ...and the Cross-Section of Expected Returns

Forty years ago, one of the first tests of the Capital Asset Pricing Model (CAPM) found that the market beta was a significant explanator of the cross-section of expected returns. The reported t-ratio of 2.57 in Fama and MacBeth (1973) comfortably exceeded the usual cutoff of 2.0. However, since that time, hundreds of papers have tried to explain the cross-section of expected returns. Given the known number of factors that have been tried and the reasonable assumption that many more factors have been tried but did not make it to publication, the usual cutoff levels for statistical significance are not appropriate. We present a new framework that allows for multiple tests and derive recommended statistical significance levels for current research in asset pricing.

We begin with 312 papers that study cross-sectional return patterns published in a selection of journals. We provide recommended p-values from the first empirical tests in 1967 through to present day. We also project minimum t-ratios through 2032 assuming the rate of "factor production" remains similar to the recent experience. We present a taxonomy of historical factors as well as definitions.[3]

[3]We also provide a link to a file with full references and hyperlinks to the original articles: *http://faculty.fuqua.duke.edu/˜charvey/Factor-List.xlsx.*

Our research is related to a recent paper by McLean and Pontiff (2013) who argue that certain stock market anomalies are less anomalous after being published.[4] Their paper tests the statistical biases emphasized in Leamer (1978), Ross (1989), Lo and MacKinlay (1990), Fama (1991) and Schwert (2003).

Our paper also adds to the recent literature on biases and inefficiencies in cross-sectional regression studies. Lewellen, Nagel and Shanken (2010) critique the usual practice of using cross-sectional R^2s and pricing errors to judge the success of a work and show that the explanatory powers of many previously documented factors are spurious.[5] Balduzzi and Robotti (2008) challenge the traditional approach of estimating factor risk premia via cross-sectional regressions and advocate a factor projection approach. Our work focuses on evaluating the statistical significance of a factor given the previous tests on other factors. Our goal is to use a multiple testing framework to both re-evaluate past research and to provide a new benchmark for current and future research.

We tackle multiple hypothesis testing from the frequentist perspective. Bayesian approaches on multiple testing and variable selection also exist. However, the high dimensionality of the problem combined with the fact that we do not observe all the factors that have been tried poses a big challenge for Bayesian methods. While we propose a frequentist approach to overcome this missing data issue, it is unclear how to do this in the Bayesian framework. Nonetheless, we provide a detailed discussion of Bayesian methods in the paper.

There are limitations to our framework. First, should all factor discoveries be treated equally? We think no. A factor derived from a theory should have a lower hurdle than a factor discovered from a purely empirical exercise. Neverthe-

[4] Other recent papers that systematically study the cross-sectional return patterns include Subrahmanyam (2010), Green, Hand and Zhang (2012, 2013).

[5] A related work by Daniel and Titman (2012) constructs more powerful statistical tests and rejects several recently proposed factor models.

less, whether suggested by theory or empirical work, a t-ratio of 2.0 is too low. Second, our tests focus on unconditional tests. It is possible that a particular factor is very important in certain economic environments and not important in other environments. The unconditional test might conclude the factor is marginal. These two caveats need to be taken into account when using our recommended significance levels for current asset pricing research.

While our focus is on the cross-section of equity returns, our message applies to many different areas of finance. For instance, Frank and Goyal (2009) investigate around 30 variables that have been documented to explain capital structure decisions of public firms. Welch and Goyal (2004) examine the performance of a dozen variables that have been shown to predict market excess returns. These two applications are ideal settings to employ multiple testing methods.[6]

1.3 Multiple Testing in Financial Economics

In many areas of economic research, answers to broad research questions emerge through time via various studies. For example, in development economics, scores of papers have tried to explain why countries grow at different rates.[7] In financial economics, hundreds to factors have been proposed to explain the cross-section of stock returns.[8] Similar to the meta-analysis approach in medical research, it is often useful to compare results from different studies. Multiple hypothesis testing is an essential tool for such an exercise. We propose a new method of multiple testing that accounts for both correlation of the tests and the fact that the meta-researcher is faced with limited data — often in the form of test statistics from previously published papers.

[6]Harvey and Liu (2014a) show how to adjust Sharpe Ratios used in performance evaluation for multiple tests.

[7]See Petrakos et. al (2007) for a recent review on economic growth.

[8]See Harvey, Liu and Zhu (2013b), McLean and Pontiff (2013) and Subrahmanyam (2010).

When testing multiple hypotheses, the primary challenge is to guard against false positive results. The statistics literature has established many criteria of Type I error rates in the context of multiple testing. Two commonly used criteria are the family-wise error rate (FWER) and the false discovery rate (FDR).[9] FWER is the probability of making at least one false discovery. FDR is the expected proportion of falsely rejected hypotheses. Many alternative definitions of error rates exist that try to modify the traditional FDR. These error rate definitions are useful in that they help us evaluate the accept/reject decisions for a set of hypothesis tests as a whole.

Several methods have been proposed to control the various error rates. The most familiar methods for controlling FWER are the Bonferroni correction and the Holm (1979) correction. Benjamini and Hochberg (1995) propose a step-down procedure to control FDR under independent test statistics. Benjamini and Yekutieli (2001) show that the Benjamini-Hochberg procedure also controls FDR under certain dependence structure. They also propose a modified procedure that controls FDR under arbitrary dependence structure. These methods are simple to use as they only involve the summary t-statistics or p-values for each individual hypothesis test. However, they are designed to control FWER or FDR under a general data structure and are often too conservative (i.e., too few true discoveries) for the particular data at hand.

Another strand of literature suggests a bootstrap based permutation approach.[10] The permutation test resamples the entire dataset several times and constructs an empirical distribution for the pool of test statistics. The empirical distribution then serves as the reference distribution for determining the cutoff values. This approach incorporates the data structure, in particular the correlation structure among the individual test statistics. Therefore, it is less conservative than the aforementioned

[9] Holm (1979) is the first to formally define the family-wise error rate (FWER). Benjamini and Hochberg (1995) define the false discovery rate (FDR).

[10] See Westfall and Young (1993) and Ge et al. (2003).

methods. However, permutation tests are computationally challenging. Moreover, permutation tests also require the knowledge of each individual dataset based on which the t-statistic or p-value constructed. In cases when this information is not available, permutation tests are not feasible.

Unfortunately, many interesting empirical inquiries in economics and science do not align with both the simple adjustment (e.g., Bonferroni and Benjamini-Hochberg procedure) and the permutation resampling approach. First, many economic variables are influenced by common shocks and are, hence, correlated. This means simple adjustment like Bonferroni or Benjamini-Hochberg are overly conservative and misleading as independence among test statistics is unrealistic. Ideally, correlation should be modeled so that its impact on multiple testing can be evaluated. Second, when a collection of previous studies are pooled together, we often do not have the luxury of having the original dataset for each study. All we have is the single test statistic that summarizes the significance a research finding. For example, Harvey, Liu and Zhu (2013b) study more than 300 factors that purportedly explain the cross-section of stock returns. It would be infeasible for them to rebuild each of the original datasets. In this case, permutation tests cannot be used.

We propose a new framework that makes it easy to evaluate the impact of correlation while at the same time only using summary statistics. We start by modeling the distribution of null hypotheses. Motivated by standard Bayesian hypothesis testing, we use a parametric mixture distribution to succinctly capture how null and non-null hypotheses are drawn. Next, we decompose commonly used test statistics into the sum of score statistics and use the Pearson correlation among the contemporaneous score statistics to model the dependence among test statistics. Then, we estimate model parameters by matching key moments of model implied and observed summary statistics. Finally, we find the threshold levels for hypothesis testing by equaling the exactly calculated Type I error rate under the estimated parameter values to a pre-specified significance level.

Our method achieves similar goals as many methods in meta-analysis in medical research. In meta-analysis, researchers combine samples of multiple studies to gain statistical power in detecting subtle effects. Recent advances in meta-analysis[11] make it possible to combine results even without individual-sample data. Our method is different from these studies in an important way. The individual studies in meta-analysis usually have a common null (e.g., a gene type does not affect a certain trait) and alternative hypothesis (e.g., how much a gene type affects a certain trait). In other words, the joint null and alternative hypothesis in multiple testing is exactly the same as the null and alternative hypothesis for each individual test. This makes the calculation of Type I error rate straightforward as everything is conditional on the common null being true. In our context, although under the cover of a broad question, different studies are testing different hypotheses. As a result, the composite alternative hypothesis is much more complicated than in traditional meta-analyses. This makes the calculation of certain Type I error rate (e.g., FDR) impossible using the prevalent methods. We overcome this problem by postulating a probabilistic model for the set of null hypotheses.

1.4 Organization

This dissertation is organized as follows. Chapter 2 develops the theoretical framework of generalized entropy bounds and relates to existing nonparametric bounds. It then makes inference on disaster models using bounds implied by index option returns. Chapter 3 follows the insights of Chapter 2 to study dynamic models. It incorporates model implied state variables as conditioning information to further distinguish candidate models. Chapter 4 first documents more than 300 risk factors that have been discovered by the academic literature. It then introduces a multiple

[11] See Conneely and Boehnke (2007, 2010).

testing framework to adjust the t-ratios of these factors. It also proposes a new multiple testing framework that can simultaneously model test correlation and hidden tests. Chapter 5 studies this new framework in greater detail through a simulation study.

2

Index Option Returns and Generalized Entropy Bounds

2.1 A unifying theory on non-parametric bounds

Hansen and Jaganathan (1991), Snow (1991), Bansal and Lehmann (1997) and Alvarez and Jermann (2005) derive non-parametric bounds under the basic no-arbitrage condition. Depending on the forms of the non-linear transformations of the pricing kernel, strong no-arbitrage condition may be required to generate meaningful bounds. For instance, the entropy bound employs a logarithmic transformation of the pricing kernel. As a result, it only makes sense if the pricing kernel is strictly positive with probability one. To the contrary, the variance bound by Hansen and Jaganathan (1991) in general has no sign restrictions [1] since the pricing kernel is raised to the second power. To be specific about the asset pricing environment and facilitate discussion, I briefly introduce some notations that will be used throughout the paper.

Let $\aleph$ be the collection of gross returns. Conceptually, it includes returns of all tradable assets and portfolios of them. Under the assumption of no-arbitrage, there

[1] Hansen and Jaganathan (1991, 1994) consider generalizations for which the pricing kernel is restricted to be nonnegative or strictly positive.

exists a pricing kernel M that prices all returns in $\aleph$, i.e.,

$$E[MR] = 1, \forall R \in \aleph. \tag{2.1}$$

Hansen and Jaganathan (1991, 1994) define Q^{++} and Q^{+} to be the set of strictly positive and nonnegative pricing kernels, respectively. Similarly, I define $\aleph^{++} = \{R : R \in \aleph \text{ and } R > 0 \text{ with probability one}\}$ and $\aleph^{+} = \{R : R \in \aleph \text{ and } R \geqslant 0 \text{ with probability one}\}$. For the same reason as in the entropy bound, I generally require $M \in Q^{++}$ and $R \in \aleph^{++}$ to produce meaningful bounds. Therefore, except for some discussions on weaker conditions towards the end, I impose these two constraints for the rest of this section. Notice that $M \in Q^{++}$ is an implication of the strong no-arbitrage condition and $R \in \aleph^{++}$ is a weak condition for gross returns of primitive assets due to limited liability. However, a portfolio of assets with excessive short positions can generate negative returns with a positive probability. The theories I develop will not apply to these portfolio strategies.

2.1.1 A continuum of new bounds

Let $M \in Q^{++}$ and $R \in \aleph^{++}$ be the stochastic discount factor and an arbitrary return, respectively. Under the no-arbitrage condition, we have the following proposition:

Proposition 1. *: $E(M^{\frac{1}{p}}) \leqslant [E(R^{-\frac{q}{p}})]^{\frac{1}{q}}$, for any $p > 1, q > 1, \frac{1}{p} + \frac{1}{q} = 1$.*

Proof. The proof involves simple manipulations of the Euler equation and Hölder's inequality.

$$\begin{aligned} E(M^{\frac{1}{p}}) &= E[(MR)^{\frac{1}{p}} R^{-\frac{1}{p}}] \\ &\leqslant [E([(MR)^{\frac{1}{p}}]^{p})]^{\frac{1}{p}} \cdot [E(R^{-\frac{1}{p}})^{q}]^{\frac{1}{q}} \\ &= [E(MR)]^{\frac{1}{p}} [E(R^{-\frac{q}{p}})]^{\frac{1}{q}} \\ &= [E(R^{-\frac{q}{p}})]^{\frac{1}{q}}. \end{aligned}$$

The second line applies Hölder's inequality to $(MR)^{\frac{1}{p}}$ and $R^{-\frac{1}{p}}$, and the last line uses the Euler equation $E(MR) = 1$. $\square$

The above proof is distinctly different from the proof of the HJ variance bound or, more generally, Snow's high-moment bounds. These bounds place restrictions on the p-th moment of the pricing kernel, with p greater than one. As a result, direct applications of Cauchy-Schwarz inequality or Hölder's inequality on the no-arbitrage condition suffice[2]. For fractional powers on the pricing kernel as in Proposition 1, the trick is to first create a power-transformed gross return to go with the pricing kernel and then annihilate them both through the no-arbitrage condition by applying Hölder's inequality in the middle.

Proposition 1 bounds the $1/p$-th moment of M by the $-q/p$-th moment of a return. As p runs from one to $+\infty$, $1/p$ covers every value in $(0,1)$. At the same time, $-q/p = 1/(1-p)$ goes from $-\infty$ to zero. Therefore, we are exhausting negative moments of the return on the right hand side. However, due to the symmetry of M and R in the no-arbitrage condition, we can obtain a continuum of bounds on negative moments of the pricing kernel by switching M and R in Proposition 1.

Corollary 2. : $E(M^{\delta}) \geqslant [E(R^{\frac{-\delta}{1-\delta}})]^{1-\delta}$, $\forall \delta \in (-\infty, 0)$.

We rewrite the bounds in Proposition 1 to make them conformable with the notations in Corollary 1:

$$E(M^{\delta}) \leqslant [E(R^{\frac{-\delta}{1-\delta}})]^{1-\delta}, \forall \delta \in (0,1). \tag{2.2}$$

Combining Corollary 1 and equation (2.2), we find lower bounds on $E(M^{\delta})$ when $\delta < 0$ and upper bounds when $\delta \in (0,1)$. The change of direction at zero seems cumbersome and inevitable, but I will show later that it is simply a matter of scaling. Under appropriate transformations, the system of bounds will be smoothly connected

[2]For an introduction on Hölder's inequality, see Casella and Berger (2001), Chap.4.

at zero.

The bounds developed above apply for any $R \in \aleph^{++}$. To provide the tightest restrictions on the pricing kernel, we can search for the optimal return R corresponding to each power δ. This is similar to the search procedure in Snow (1991) and Bansal and Lehmann (1997). In particular, define $\rho(\delta)$ as

$$\rho(\delta) = \begin{cases} \sup_{R \in \aleph^{++}} [E(R^{\frac{-\delta}{1-\delta}})]^{1-\delta} & \text{if } \delta \in (-\infty, 0), \\ \\ \inf_{R \in \aleph^{++}} [E(R^{\frac{-\delta}{1-\delta}})]^{1-\delta} & \text{if } \delta \in (0, 1). \end{cases} \tag{2.3}$$

Then $\rho(\delta)$ gives the sharpest lower (upper) bound on $E(M^{\delta})$ when $\delta \in (-\infty, 0)$ ($\delta \in (0, 1)$).

2.1.2 Interpreting bounds

What are the economic stories behind these bounds? In particular, what do the two sides of these inequalities measure? Do equalities reveal something fundamental about the economy? I provide a utility-based interpretation of my bounds.

Let us first introduce a risk-aversion index $\gamma(\delta)$ defined as

$$\gamma(\delta) \equiv \frac{1}{1-\delta}, \;\; \delta \in (-\infty, 0) \cup (0, 1).$$

Note that $\gamma(\delta)$ has a well-defined support as a risk-aversion coefficient: $\gamma(\delta) \in (0, 1)$ if $\delta \in (-\infty, 0)$ and $\gamma(\delta) \in (1, +\infty)$ if $\delta \in (0, 1)$. Next, define the *augmented return space* as

$$\aleph^{**} = \{R : E(MR) = 1 \text{ and } R > 0 \text{ with probability one}\}.$$

It is crucial to see the difference between $\aleph^{++}$ and $\aleph^{**}$: the former contains returns of assets that are tradable in the market while the latter contains all positive returns that satisfy the no-arbitrage condition. In other words, $\aleph^{++}$ includes whatever the market has while the potentially much larger $\aleph^{**}$ includes what the market could have. The difference between $\aleph^{++}$ and $\aleph^{**}$ measures the degree of market completeness.

Endowed with this augmented return space $\aleph^{**}$, I seek to solve the portfolio choice problem for an agent with unit endowment and risk-aversion $\gamma(\delta)$. This optimization problem can be written as

$$U_\delta(M) \quad = \quad \sup_{R\in\aleph^{**}} E[\frac{R^{1-\gamma(\delta)}}{1-\gamma(\delta)}]. \tag{2.4}$$

The maximized utility $U_\delta(M)$ depends on the discount factor M, whose information is already embedded in $\aleph^{**}$. The following proposition gives the solution to this maximization problem.

Proposition 3. *The solution to the maximization problem in (2.4) is given by*

$$U_\delta(M) \quad = \quad \frac{[E(M^{\frac{\gamma(\delta)-1}{\gamma(\delta)}})]^{\gamma(\delta)}}{1-\gamma(\delta)} = \frac{[E(M^\delta)]^{\frac{1}{1-\delta}}}{1-\gamma(\delta)}, \tag{2.5}$$

$$\tilde{R}_\delta(M) \quad = \quad M^{-\frac{1}{\gamma(\delta)}}/E(M^{\frac{\gamma(\delta)-1}{\gamma(\delta)}}). \tag{2.6}$$

Proof. The appendix contains a detailed proof. The inequalities in Proposition 1 and Corollary 1 establish the finiteness of the objective function, making the optimization problem well-defined. The proof then proceeds in two steps. First, the optimal portfolio choice $\tilde{R}_\delta$ is solved as a function of the Lagrange multiplier associated with $E(MR) = 1$, viewed as a budget constraint. Second, the Lagrange multiplier itself is solved using the no-arbitrage condition. $\square$

Now the economic meanings of bounds stand out. To ease interpretation, we can rewrite the bounds shown in Corollary 1 and equation (2.2) as

$$\frac{[E(M^\delta)]^{\frac{1}{1-\delta}}}{1-\gamma(\delta)} \geqslant \frac{E[R^{1-\gamma(\delta)}]}{1-\gamma(\delta)}, \forall R \in \aleph^{++}. \tag{2.7}$$

By Proposition 2, the quantity on the left hand side is the maximized utility over the augmented return space $\aleph^{**}$ for an agent with a risk-aversion coefficient of $\gamma(\delta)$. It is the highest achievable utility if the market is complete in the sense that $\aleph^{**} = \aleph^{++}$ or, as a weaker requirement, the optimal choice $\tilde{R}_\delta(M)$ is actually tradable, i.e., $\tilde{R}_\delta(M) \in \aleph^{++}$. For a given risk aversion $\gamma(\delta)$, the log of $\tilde{R}_\delta(M)$ loads negatively on the log pricing kernel, i.e.,

$$\log \tilde{R}_\delta(M) = A_\delta(M) - \frac{1}{\gamma(\delta)} \log M, \tag{2.8}$$

where $A_\delta(M)$ is a scaling constant. Given the stylized fact that market moves counter-cyclically, my bounds are easier to satisfy at equality and can potentially be more informative than HJ or Snow's high-moment bounds.

From an asset pricing perspective, although the marginal investor determines the discount factor, investors with different levels of risk aversion all have a say in the behavior of the discount factor. Their optimal portfolio choices automatically place a sequence of thresholds that the discount factor has to overcome. As the power δ goes through its admissible region, we are essentially running through the support of the risk-aversion coefficient of power-utility agents. This interpretation is in spirit similar to Bansal and Lehmann (1997)'s interpretation of the growth-optimal portfolio, albeit I am able to significantly generalize their argument.

From a methodological perspective, it is interesting to compare the approach I have taken to interpret my bounds and that of HJ (See Hansen and Jaganathan, 1991, Gallant, Hansen and Tauchen, 1990 and Bekaert and Liu, 2003). HJ bounds are constructed by projecting the pricing kernel onto the space of available asset payoffs. The L_2-norm of the projected pricing kernel has the minimal standard deviation across all valid pricing kernels. I start from a candidate pricing kernel and ask what an optimizing agent will do in an ideal world where all "admissible" returns are tradable. Consequently, by limiting the asset space to tradable asset returns, the agent's real-world objective function dictates a lower bar that the starting candidate pricing kernel has to satisfy. As a matter of fact, in a strict duality sense, these two approaches are complementary to each other. I rigorously define the duality concept and prove it in the appendix. On a practical level, HJ's approach is more transparent when certain moment of the pricing kernel (e.g., variance, sharpe ratio, etc.) is the focus and my approach is more intuitive when a return moment (e.g., a CRRA investor's objective function) is the interest.

2.1.3 Characterizing the non-parametric bound universe

Thus far, I have established bounds for various moments of the pricing kernel, with zero being the only undefined case. Additionally, I extend the log-utility interpretation of the entropy bound to general power utilities. These results prompt us to wonder if the entropy bound is the one that fills the hole of my bound spectrum. Indeed, it is. The following proposition formally establishes this.

Proposition 4. *The bounds given in Proposition 1 and Corollary 1 both imply the entropy bound:* $E(\log(M)) \leqslant -E(\log(R))$.

Proof. The proof applies the same intuition as how power utility converges to log-utility. I only show how bounds in Proposition 1 imply the entropy bound. Essentially the same proof can be done for bounds in Corollary 1. I start by scaling the bounds in equation (2.2):

$$\frac{E(M^{\delta}) - 1}{\delta} \leqslant \frac{[E(R^{\frac{-\delta}{1-\delta}})]^{1-\delta} - 1}{\delta}.$$

This is true because $\delta > 0$. Taking limits as $\delta \downarrow 0$ and under regularity conditions,[3] the left hand side is readily seen to converge to $E(\log(M))$ using L'Hôspital's rule.

[3] We need conditions on moments of M and R to be able to exchange limits and expectation. Dominated convergence will suffice. See Davidson (1994) Part IV for some specific conditions.

A careful application of the rule to the right hand side will also deliver $-E(\log(R))$ as the limit. $\square$

We are now ready to summarize the non-parametric bound universe that the literature has discovered. Figure 2.1 shows a diagram of this bound universe. When the power equals one, the expected marginal rate of substitution is bounded within $[\frac{1}{\min R}, \frac{1}{\max R}]$ for a generic $R \in \aleph^{++}$. This seemingly informative bound becomes redundant in the presence of a risk-free rate R_f, since $E(M) = 1/R_f$. Starting from $s = 1$ and going right, one encounters the spectrum of Snow's high-moment bounds and HJ bound is sitting at $s = 2$. Going left, one sees the continuum of bounds I just developed, and the BL/AJ entropy bound fills the hole at $s = 0$. It is intriguing to see the symmetric pattern of these bounds around $s = 1$, particularly in light of the order by which they are discovered by the literature.[4]

Lastly, we ask if the bound system is complete. This is more than a technical question because we do not want to leave out any information on the pricing kernel that can be assessed by the asset market. In particular, given the existence of these one-sided inequalities for essentially any moment of the pricing kernel, one may wonder whether other bounds, possibly with opposite directions of inequalities, can further enrich the bound universe. After all, a two-sided bound certainly looks

[4]Recent papers by Almeida and Garcia (2012,2013) proposes similar nonparametric bounds. However, there are important differences between their works and my work. First, they follow the Hansen and Jaganathan (1991) approach but use a new objective function to derive bounds. One can insert their optimal solution into the objective function to obtain a bound that is similar to mine. However, these bounds are not exactly the same as their bound assumes a known risk-free rate. Moreover, I propose a framework that is different than the Hansen and Jaganathan (1991) approach to systematically construct and understand nonparametric bounds. It highlights the optimizing behavior of the economic agents, which is new to the literature. The resulting bounds are also in their cleanest forms, i.e., all statistical uncertainty and discrete time approximations are passed on to the right-hand side of the bound through a portfolio optimization problem. Second, I discuss my bounds in comparison with the basic entropy bound. Through a moment-expansion exercise, I analytically show how my bounds capture higher-order moments of both the pricing kernel and asset returns. These insights would be difficult to see if the bounds involve sample moments of returns. Third, I use index option returns to confront state-of-the-art asset pricing models. This exercise highlights the power of bounds and, in my view, is most relevant application of my bounds.

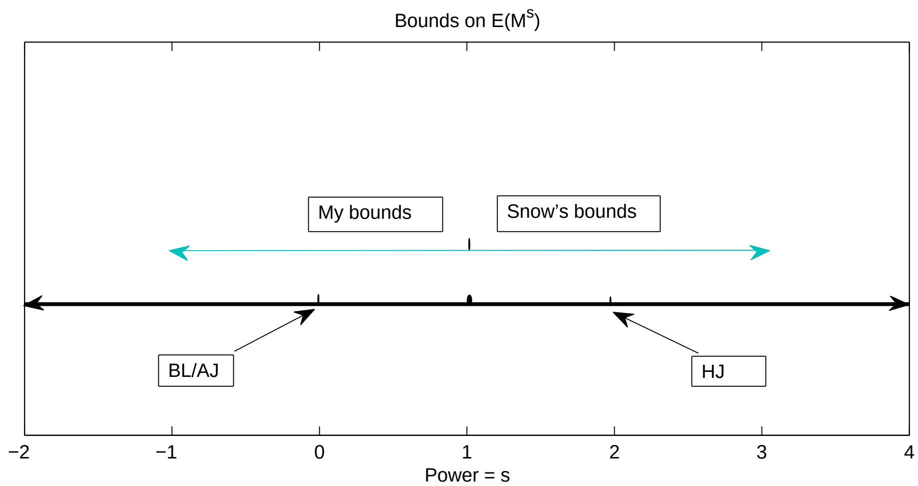


FIGURE 2.1: **The non-parametric bound universe**

more appealing than a one-sided bound as both sides are constrained. The following proposition eliminates such possibilities and indirectly shows the exhaustiveness of the above bound universe.

Proposition 5. *For a given power s and the corresponding upper (lower) bound on $E(M^s)$, the lower (upper) side of $E(M^s)$ is generally unbounded. Hence, the non-parametric bound system is exhaustive.*

Proof. The idea is to construct a sequence of pricing kernels that can all price a return but has unbounded limit for a given moment. I leave this proof to the appendix. □

2.1.4 Discussions and extensions

The new continuum of bounds can be extended along several dimensions. First, it can be adapted to study the dynamic behavior of the pricing kernel. Let $M_{t,t+n} = M_{t+1}M_{t+2}\dots M_{t+n}$ be the time aggregated pricing kernel and $R_{t,t+n} = R_{t+1}R_{t+2}\dots R_{t+n}$ be a generic multi-period return. The long-horizon asset returns provide bounds on the unconditional moments of the time aggregated pricing kernel. These unconditional moments of the multi-period pricing kernel reveal the dynamic dependency of the single period pricing kernels. Different moments shed lights on different forms of dynamic dependency. For instance, one natural way to scale an n-period pricing kernel is to take the fractional power $1/m$ on the time aggregated pricing kernel. A bound on thus scaled kernel is given by

$$E(M_{t,t+n}^{\frac{1}{m}}) \leqslant [E(R_{t,t+n}^{\frac{-1}{m-1}})]^{\frac{m}{m-1}}, m \geqslant 2.$$

When $m \to \infty$, properly scaled version of the above bound converges to the multi-period entropy bound. Backus, Chernov and Zin (2011) use the multi-period entropy bound to study time-dependency in discount rates. Setting m at n, the left hand side becomes $E[\exp(\frac{1}{n}\sum_{j=1}^{n}\log M_{t+j})]$ so the average of the log pricing kernel is revealed. By decomposing the pricing kernel into a permanent and a transitory component (Alvarez and Jermann, 2005), such scaling leaves the expectation of the permanent component intact while allows the transitory component to decay at a rate of n. Bounds give us information on how fast they decay. Liu (2013) builds on this insight to diagnose dynamic asset pricing models using the generalized entropy bounds.

Second, conditioning information can be incorporated to sharpen bounds on unconditional moments (Bekaert and Liu, 2004, Gallant, Hansen and Tauchen, 1990 and Ferson and Siegel, 2003). Notice that simply adding instruments to the conditional Euler equation (Hansen and Jaganathan, 1990) is different from using returns that are generated from a dynamic trading strategy. The utility-based interpretation of my bounds naturally favors the latter approach. As shown in Ferson and Siegel (2003), the use of returns of dynamic strategies significantly sharpens HJ bounds. Liu (2013) uses conditioning information to differentiate key predictability assumptions in leading asset pricing models.

2.2 Bound informativeness

Having established a system of bounds, one may wonder what unique insights can a bound on a certain moment of the discount factor give us. After all, the exploration of a continuum of powers is physically impossible so we would like to select a few bounds that are both representative and informative. With this goal in mind, I perform a dissection of the bound universe. Much like a doctor performing surgery, I need a "surgical knife" to decompose the bound system. I first develop a useful tool in studying bounds. I define a quantity that is a natural generalization of the entropy concept popularized by Bansal and Lehmann (1997), Alvarez and Jermann (2005) and Backus, Chernov and Zin (2011) of the recent asset pricing literature. I show that it is both economically meaningful and analytically tractable. With this tool in hand, I apply the cumulant-expansion technique (Backus, Chernov and Martin, 2011 and Martin, 2008) to examine both sides of a bound. Lastly, a concrete example is given to gain deeper understanding of the workings of a bound.

2.2.1 A useful quantity

The recent asset pricing literature proposes a convenient measure to study the link between the pricing kernel and asset returns (Alvarez and Jermann, 2005, Bansal and Lehmann, 1997, Backus, Chernov and Zin, 2011 and Martin, 2011).[5] The entropy of the marginal rate of substitution is used to measure pricing kernel dispersion and is bounded below by the continuously compounded risk premium:

$$L(M) \equiv \log E(M) - E(\log M) \geqslant E(\log R) - \log(R_f), \tag{2.9}$$

where $R \in \aleph^{++}$ is an arbitrary return and $R_f = 1/E(M)$ is the gross risk-free rate, assuming one exists. Researchers rely on the entropy bound to gauge the amount of dispersion that an asset pricing model has to generate. However, considering the continuum of bounds that are relatives of the entropy bound, we would expect to gain additional insights by using other bounds. I propose a quantity that is a natural

[5]Notably, entropy is gaining popularity in many fields of economics and finance. See Stutzer (1996), Hansen and Sargent (2008), Ghosh, Julliard and Taylor (2011) and Van Nieuwerburgh and Veldkamp (2010).

generalization of the original entropy concept. The continuum of bounds developed in the previous section can then be brought in to further restrict the pricing kernel. In essence, I am normalizing the system of bounds in reference to the entropy bound.

The *Generalized Entropy Function* (GEF) of a positive pricing kernel is defined as:

$$GEF(s;M) \equiv \log E(M) - \frac{1}{s}\log E(M^s) \tag{2.10}$$

for any real-valued s. It is an extension of the original entropy because its limit at zero is exactly the entropy, i.e.,

$$\lim_{s\to 0} GEF(s;M) = L(M).$$

Assuming the finiteness of all moments, $GEF(s;M)$ is an everywhere continuous function on the real line. Moreover, many convenient properties of entropy are maintained by the GEF. For instance, GEF equals zero at a power s if and only if M is a constant. Similar to entropy, it is scale-invariant, i.e., $GEF(ws;M) = GEF(s;M)$ for a constant w. Hence, GEF leaves the pricing kernel numeraire invariant. This is an appealing property empirically because we do not need to worry about the adjustment between a nominal and a real pricing kernel. Additionally, GEF is pivotal around $(1,0)$ in the sense that every GEF has to pass $(1,0)$ on the two-dimensional plane. This gives a fixation point to anchor the GEF's corresponding to different pricing kernels. Finally, in the familiar lognormal case, $GEF(s;M) = (1-s)\sigma_M^2/2$ where σ_M^2 is the variance of the log pricing kernel.

The bound universe developed in the previous section can be brought in to restrict $GEF(s;M)$. The implied restrictions can be shown as:

$$GEF(s;M) \geqslant \frac{s-1}{s}\log E(R^{\frac{s}{s-1}}) - \log(R_f), \forall s \in (-\infty, 1). \tag{2.11}$$

Notice how the two types of bounds in Corollary 1 and equation (2.2) nicely line up with each other in terms of the direction of inequalities. The undesirable flip in direction at zero disappears once we introduce the generalized entropy function. From this perspective, GEF seems to be a more appropriate apparatus in studying $E(M^s)$ when $s < 1$. When $s > 1$, Snow's continuum of high-moment bounds imply

$$GEF(s; M) \leqslant \frac{s-1}{s} \log E(R^{\frac{s}{s-1}}) - \log(R_f), \forall s \in (1, +\infty). \qquad (2.12)$$

Figure 2.2 plots a generic GEF with asset market bounds.

Of course, sharper restrictions can be found by explicitly searching for the optimal bounds, as in equation (2.3). Given the convenience offered by GEF, for the rest of the paper I will focus on bounds given in the form of (2.11) or (2.12) unless otherwise specified. I will refer to the system of bounds given in (2.11) as the generalized entropy bounds or, with a slight abuse of terminology, simply entropy bounds. The bounds in (2.12) are termed high-moment bounds.

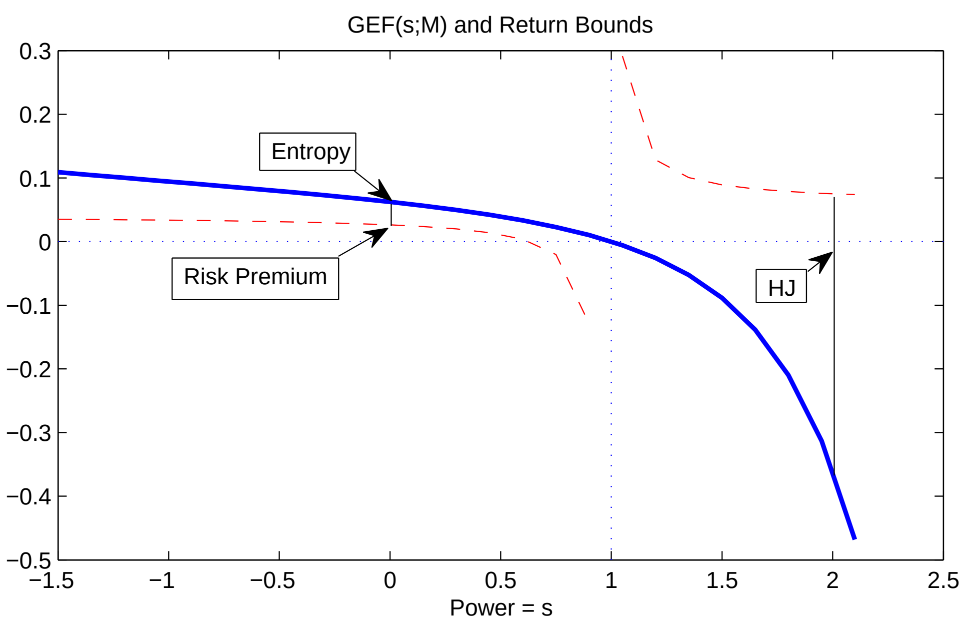


FIGURE 2.2: **A typical plot of $GEF(s;M)$ and asset market bounds.** This figure plots a generic GEF and asset market bounds. The thick solid line depicts the GEF and the thin dashed line depicts asset market bounds.

2.2.2 *Expanding the GEF*

The Cumulant-Generating Function (CGF) is another recently developed tool to study higher order moments of the pricing kernel (Backus, Chernov and Martin, 2011 and Martin, 2008). By Taylor-expanding the log expected pricing kernel into a power series, it shows how higher order moments contribute to the overall entropy. Backus, Chernov and Martin (2011) use entropy as a measure of dispersion and study how a disaster model and an empirical option pricing model imply different moment characteristics of the pricing kernel. Assuming representative agent and i.i.d. consumption growth rates, Martin (2008) links fundamentals to moments of consumption growth and performs a calibration exercise. I contribute to this literature by showing that asset returns provide valuable information about the entire CGF and not just at zero, which corresponds to the original entropy. This significantly strengthens the link between asset pricing models and market returns and can potentially help us better distinguish candidate models.

I start by performing a Taylor expansion of our newly defined $GEF(s;M)$. This amounts to Taylor expanding $E(M^s) = E(e^{s \log M})$ around $s = 0$:

$$\begin{aligned} GEF(s;M) &= \sum_{i=1}^{\infty} \frac{\kappa_i(\log M)}{i!} - \frac{1}{s} \sum_{i=1}^{\infty} \frac{\kappa_i(\log M)}{i!} s^i \\ &= \sum_{i=2}^{\infty} \frac{\kappa_i(\log M)}{i!} (1 - s^{i-1}) \\ &= \frac{\kappa_2(\log M)}{2!}(1-s) + \frac{\kappa_3(\log M)}{3!}(1-s^2) \\ &\quad + \frac{\kappa_4(\log M)}{4!}(1-s^3) + \frac{\kappa_5(\log M)}{5!}(1-s^4) \ldots . \end{aligned} \tag{2.13}$$

The first two lines Taylor-expand the two components in $GEF(s;M)$ and group similar terms. The last line explicitly writes out the first few terms in this expansion. Here $\kappa_i(\log M)$ denotes the i-th "cumulant" of the log discount factor and is defined as the i-th derivative of $\log E(e^{s \log M})$ at $s = 0$. Cumulants are closely related to moments: $\kappa_1(\log M)$ and $\kappa_2(\log M)$ are the mean and variance of $\log M$, respectively and $\kappa_3(\log M)$ and $\kappa_4(\log M)$ are related to the usual skewness (ν_1) and excess kurtosis (ν_2) through: $\nu_1 = \kappa_3(\log M)/[\kappa_2(\log M)]^{\frac{3}{2}}$ and $\nu_2 = \kappa_4(\log M)/[\kappa_2(\log M)]^2$ (Backus, Chernov and Martin, 2011).

The expansion of GEF reveals that cumulants are weighted by polynomials of s. In particular, the i-th scaled cumulant $\kappa_i(\log_M)/i!$ is multiplied by $(1 - s^{i-1})$. By varying the value of the argument s, $GEF(s; M)$ puts different weights on different moments. In this way, $GEF(s; M)$ conveys information about all the moments of the pricing kernel. In particular, when evaluated at $s = 0$, $GEF(s; M)$ equals the original entropy, which is an overall sum of $\{\kappa_i(\log M)/i!\}_{i=1}^{\infty}$.

I argue that $GEF(s; M)$ is especially useful in teasing out the tail information of the pricing kernel. Take, for example, a standard disaster model along the lines of Barro (2006, 2009). In such a model, large drops in consumption in disastrous states generate a huge amount of negativity in skewness and all the other odd moments. Consequently, the marginal rate of substitution, which loads negatively on consumption growth, will display excess positivity for all odd moments. At the same time, even moments will mechanically increase as well with the presence of outliers in state prices. This creates an identification problem for the ultimate source of dispersion. Backus, Chernov and Marin (2011) argue that odd cumulants in the original entropy expansion reflect the inherent asymmetry in jumps. However, given that all moments are equally weighted at $s = 0$, it is difficult to see the differential effect between odd and even moments. I suggest taking large negative s values to inflate this wedge. Large negative s makes the loadings associated with even moments positive and those associated with odd moments negative. Thus, a "net" jump effect is singled out by taking the difference of odd and even moments. In fact, $s < -1$

is the only parameter region for the net jump effect to appear. This highlights the importance of my bound system developed in the previous sections.

Similar expansions can be applied to returns on the right hand side of bounds. This is important in that it gives us guidance on the selection of the most informative asset returns. I cumulant-expand the right hand side of equation (2.11) as:

$$\begin{aligned} GEF(s;M) &\geqslant \frac{s-1}{s}\sum_{i=1}^{\infty}\frac{\kappa_i(\log R)}{i!}(\frac{s}{s-1})^i - \log R_f \\ &= [E(\log R) - \log R_f] + \sum_{i=2}^{\infty}\frac{\kappa_i(\log R)}{i!}(\frac{s}{s-1})^{i-1}, \end{aligned} \tag{2.14}$$

where $\kappa_i(\log R)$ is the i-th return cumulant. We see that in addition to the continuously-compounded equity risk premium $[E(\log R) - \log R_f]$, an extra term $\sum_{i=2}^{\infty}\frac{\kappa_i(\log R)}{i!}(\frac{s}{s-1})^i$ comes out of the expansion. For large negative s values, $\frac{s}{s-1}$ is close to one, so the first few higher order cumulants will enter significantly into the right hand side. This means that to search for the tightest lower bounds, we need returns that possess excess (positive) skewness and kurtosis. Option strategy returns fit these descriptions (Coval and Shumway, 2001 and Broadie, Chernov and Johannes, 2008). In the empirical section of the paper, I follow this intuition to explore the restrictions that option strategy returns place on the discount factor.

2.2.3 A complete market economy example

To better illustrate the workings of entropy bounds, I consider a complete market economy with finite states. Under this setup, limiting distributions for the discount factor can be easily derived and asset market has a transparent structure. Through this exercise, we try to gain intuition on the following questions: 1. How does the generalized entropy reduce to a tail measure when s is sufficiently negative? 2. How exactly do security returns bound the pricing kernel?

Consider a one-period economy with only two states of nature. The state prices are given by:

$$M = \begin{cases} M_1, & \text{with probability } p_1 \\ M_2, & \text{with probability } p_2 = 1 - p_1. \end{cases} \tag{2.15}$$

Imagine that state 1 is a bad state and state 2 is a normal state so that $M_1 > M_2$. Relating to a disaster model, we will have $M_1 \gg M_2$. In this economy, $GEF(s; M)$ has the following limiting behavior:

$$\begin{aligned} GEF(s; M) &= \log(EM) - \frac{1}{s}\log(EM^s) \\ &\to \log(\underbrace{p_1 \frac{M_1}{M_2} + p_2}_{\bar{W}}) \end{aligned}$$

when $s \to -\infty$.

What is $\bar{W}$? In the presence of a risk-free rate R_f, $\bar{W}$ can be expressed as $\frac{1}{R_f} \cdot \frac{1}{M_2}$. This is the expected wealth for a risk-neutral optimizing agent with an endowment of one unit of the riskless bond. To see this, notice that the expected returns for the two Arrow-Debreu securities are $1/M_1$ and $1/M_2$, respectively. A risk-neutral agent only cares about expected returns. Hence, she will shy away from the insurance asset that is relatively expensive and have a concentrated position on the second Arrow-Debreu security. Her expected end-of-period wealth is thus equal to the beginning-of-period wealth $(1/R_f)$ times the expected return of the second elementary security $(1/M_2)$. This interpretation of $\bar{W}$ reminds us of the utility-based interpretation of entropy bounds in the previous section.

Intuitively, a risk-neutral investor's portfolio choice contains superior information about tail events because her incentive in "selling" the insurance asset is the strongest. To see how the tail event probability is reflected in $\bar{W}$, note that

$$\bar{W} = \log[1 + p_1(\frac{M_1}{M_2} - 1)] \approx p_1(\frac{M_1}{M_2} - 1)$$

under mild conditions[6]. Therefore, the rare event probability p_1 is amplified by the multiplier $(\frac{M_1}{M_2} - 1)$. The more frequently tail events happen and/or the larger the state price for a disastrous state is, the more profit a risk-neutral agent can exploit. In this sense, the tail event information is partially identified through the risk neutral agent's optimization behavior.

On the other hand, the right hand side of equation (2.11) reduces to $\log(\frac{1}{R_f} \cdot E(R))$ when $s \to -\infty$. Hence, the generalized entropy bounds simply say that in expectation no return can exceed the return of the second Arrow-Debreu security. This is trivially true given the structure of the economy. What is not so trivial, even for this two-state economy, is how entropy bounds internally extract information from the discount rate and provide economically sensible links to market securities.

The example can be easily extended to an N-state case, but the marginal gain in intuition is limited. In essence, the generalized entropy bounds feed on the idea that a moment of the discount rate can be viewed as the objective function for a certain type of investors. By taking large negative moments, the rare event information stands out and this information is gauged against the maximized utility of a nearly risk-neutral agent.

2.3 Option market bounds and rare disaster models

Tail information, long recognized as a potential source to generate economic risk premiums (Rietz, 1988), has recently been elevated to quantitatively explain asset market abnormalities. Barro (2006), Barro and Ursua (2008) and Barro, Nakamura, Steinsson and Ursua (2009) extrapolate the tail distribution of the US consumption growth process by looking at international macroeconomic data. Based on the exchange economy and representative agent framework, they argue that the calibrated

[6]For a standard disaster model, in which the disaster state probability is in the range of 1%-2% and the state price ratio is within 10, the approximation is good.

rare event distribution can explain key moments of US asset returns, in particular the equity risk premium. Gabaix (2009), Wachter (2009) and Gourio (2008) extend the basic disaster model to account for other salient features of asset markets.

At the heart of the rare disaster literature is the so-called *pseudo* problem: given the rare occurrence of disasters, one cannot measure their distributions accurately based on a relatively short univariate time series. As a remedy, researchers pool data from other sources to avoid the inherent small sample problem. An alternative approach is to use asset market returns to infer the tail information in the pricing kernel. Since option prices are informative about the investors' ex-ante valuation of extreme event risks, they can be a useful source of information. Backus, Chernov and Martin (2011) use equity index options to infer the distribution of consumption growth. I also consider index options but take a different approach to study their implications on the pricing kernel. In particular, I use the newly developed nonparametric bounds to restrict the behavior of a pricing kernel. In doing so, tail information of the kernel is distilled into various moment inequalities. My approach has several advantages. First, it is based on the basic no-arbitrage condition alone and thus free from any misspecification on the linkage between the pricing kernel and asset returns. Second, no parametric model is needed to fit the cross-section of option prices. Instead, individual option trading strategies are estimated and fed into the nonparametric bounds. Lastly, a formal statistical testing framework is developed. It features the simultaneous testing of several bounds. I show the discriminatory power of the generalized entropy bounds at negative power values.

2.3.1 Data description

For the empirical analysis, I use monthly data on the S&P 500 index, the associated index options and the risk-free rate. The riskfree rate is from Kenneth French's on line data library. The full sample for index returns run from July 1926 to December 2011. The shorter sample, which coincides with the span of the option data from OptionMetrics, is from January 1996 to December 2011, yielding 192 months of data. All nominal returns are converted to ex-post real returns using the inflation rate based on CPI.

I collect European style S&P 500 index options for the period from January 1996 to December 2011 from the OptionMetrics database. The data set contains daily settlement prices for options with various strike prices and maturities, as well as liquidity measures such as open interests and trading volumes. It also includes dividend yield for the market index and interpolated zero coupon yields. They are useful for us to construct option trading strategies later on. To mitigate microstructure issues, I drop option data with average bid-ask prices less than one eighth of a dollar, with open interests less than 100 contracts or with zero trading volumes. Finally, I use put-call parity relationships to filter out data that obviously violate the no-arbitrage condition.

To construct equally-spaced monthly returns from option prices, I follow a procedure that is similar to Coval and Shumway (2001), Buraschi and Jackwerth (2001) and Driessen and Maenhout (2005). First, options with strike-to-spot ratio closest to 92%, 96% and 100% are targeted on the first trading day of each month. Next, these option contracts are followed and identified until the beginning of the subsequent month and the monthly holding period returns are calculated. In this process, option contracts that expire in the third week of the same month have to be excluded. Due to liquidity concerns, I focus on short-maturity options with around seven weeks to maturity at the buying date and about two weeks to maturity at the selling date. These options have large trading volumes and are less affected by liquidity problems (Bondarenko, 2003).

The equally-spaced return series are convenient to handle empirically, especially when we use them to confront asset pricing models. This is because most discrete time models specify a fixed sample frequency. Additionally, as argued by Driessen and Maenhout (2005), thus constructed returns are more sensitive to changes in jump or volatility risks than hold-to-maturity returns. This sensitivity is important for this paper as risk-sensitive returns can potentially provide the most informative bounds on the pricing kernel. Finally, from a more technical point of view, my bound theory requires returns to have a positive support. However, hold-to-maturity returns are at times extremely high, which creates trouble in constructing a short strategy that always generates positive returns.

Instead of using raw option returns data, I focus on returns from a few well-known derivative strategies. There are two main reasons for this: 1. These economically meaningful strategies offer clear interpretations of the sources of risks (jump risks and/or volatility risks) that are being traded; 2. The recent literature on option pricing anomalies mainly focus on these trading strategies (Bondarenko, 2003, Coval and Shuway, 2001 and Driessen and Maenhout, 2005). To be consistent and comparable with existing studies, I choose to adopt this tradition.

In particular, I take the following two strategies as the benchmark strategies:

- an out-of-the-money (OTM) put option with 96% moneyness;
- an at-the-money(ATM) straddle.

A deep OTM put is a hedge against market crashes and much less so against volatility movements. Its price is therefore only sensitive to market jump risks. On the other hand, an ATM market-neutral straddle generates profits when either the market volatility is high or when market crashes, so it is exposed to both volatility and jump risks. These two option strategies are among the most commonly traded strategies by market practitioners and have been extensively studied by the recent option pricing literature (Coval and Shumway, 2001, Jackwerth, 2000, Bondarenko, 2003), which make them the ideal choice as benchmark strategies.

In addition to the benchmark strategies, I further consider two types of "crash-neutral" variants of them. They are simply the two original options mixed with an offsetting short position in the 92%-OTM put option[7]. In doing this, large returns when market crashes are capped off for long positions in the benchmark strategies and symmetrically, short positions are protected against large downward movements of the market returns[8]. It would be interesting to see whether these alternative strategies can provide any additional information beyond what is provided by the benchmark strategies.

Table 2.1 shows the summary statistics for the returns of various derivative strategies as well as the market index. Consistent with the literature, these option strategies generate large negative average returns and Sharpe ratios. The flip side would be the potential profits generated by short positions in these strategies. Moreover, the returns are highly non-normally distributed, as reflected by the large magnitude of skewness and kurtosis. These moment characteristics of index option strategy returns will be helpful later on in forming informative bounds on the pricing kernel. Nonetheless, in spite of the magnitude, the mean returns for my sample are notably smaller and about half the size of those reported by Coval and Shumway (2001), Broadie, Chernov and Johannes (2007) and Bondarenko (2003). This is mainly driven by the instability of mean returns for these derivative strategies. The aforementioned papers mainly focus on the period before 2005 and many include the 1987 crash episode, whereas mine starts in 1996 and extends all the way to the most recent period. It is

[7]For details on the construction of the crash-neutral strategies, see Coval and Shumway (2001) and Jackwerth (2000).

[8]This is only approximately true because the beginning-of-period 92%-OTM put and 96%-OTM put may have different maturities. In fact, the deeper 92%-OTM put may have a higher price than the 96%-OTM put simply because the former's maturity is significantly longer than the latter. This creates trouble in the usual interpretation of crash-neutral strategies: for instance, a short leg in 92%-OTM put becomes a long leg. Therefore, I also create robust-crash-neutral put and straddles which essentially delete these abnormal dates. See the descriptions of Table 1 for details.

Table 2.1: **Summary statistics.** This table reports the summary statistics of the returns for the S&P 500 index and several derivative strategies. The long index series is from July 1926 to December 2011 and the short index series is from January 1996 to December 2011, which is also the time span for all the option strategy returns. The first column displays the strategy name and the last column reports the correlation of the strategy returns with the short-sample market index. "C-neutral put" and "C-neutral straddle" denote crash-neutral put and straddle returns, for which the original 96%-OTM put and ATM straddle are mixed with a short leg on 92%-OTM put option, respectively. See Coval and Shumway (2001) for the construction of the crash-neutral put and Jackwerth (2000) for the construction of the crash-neutral straddle. "R-C-neutral put" and "R-C-neutral straddle" denote robust crash-neutral put and straddle returns, respectively. They are the original crash neutral series excluding the date in which the 92%-OTM put maturity date is more than three trading weeks longer than the 96%-OTM put maturity date at the moment of buying. In doing this, six observations are deleted from the 192 monthly observations, including two months in which the 92%-OTM put has a higher price than the 96%-OTM put. Skewness and kurtosis are the standardized central third and fourth moments, respectively. The riskfree rate is 60bp annualized for the long sample and 54bp for the short sample. These rates are the inputs for the calculation of the Sharpe ratios for the corresponding samples.

	Mean	Std.	Sharp	Skew.	Kurt.	Max	Min	Corr.Index
Index,long	0.007	0.055	0.113	0.225	10.323	0.384	-0.285	NA
Index,short	0.005	0.049	0.087	-0.605	3.666	0.118	-0.177	1.000
0.92-OTM put	-0.236	0.728	-0.325	2.756	13.049	3.944	-0.877	-0.305
0.96-OTM put	-0.201	0.611	-0.330	1.519	4.839	2.261	-0.849	-0.277
1.00-OTM put	-0.082	0.627	-0.132	1.897	7.990	3.256	-0.849	-0.323
ATM straddle	-0.017	0.340	-0.052	1.702	6.809	1.501	-0.778	0.001
C-neutral put	-0.310	0.996	-0.312	-2.553	18.805	2.135	-7.438	-0.096
C-neutral straddle	-0.015	0.428	-0.035	-0.528	13.847	1.665	-2.823	0.077
R-C-neutral put	-0.212	0.724	-0.293	-0.075	7.469	3.135	-2.960	-0.119
R-C-neutral straddle	-0.025	0.422	-0.059	-0.660	14.602	2.665	-1.823	0.059

worthwhile to mention that the recent six years (2006-2011) see significant increases in returns for these strategies. For instance, the average 92%-OTM and 96%-OTM put returns are -12.8% and -13.6% per month, respectively, much larger than their sample averages in early years. A full investigation into the change in these return characteristics is beyond the scope of this paper. To the extent that my sample under-represents the true option return population and overestimates mean returns, the bounds constructed below can be regarded as conservative lower bounds on the generalized entropy function of the pricing kernel.

2.3.2 Bounds implied by option strategies

With these return data from the asset market and relying on the analytical tools developed in the previous sections, I explore their implications on the behavior of a pricing kernel. Ideally, to provide the sharpest bounds, we need to search for the optimal dynamic strategy that maximizes a certain unconditional moment of the return. However, parameter estimates for even simple static portfolio choice problems are usually very unstable, partly due to the volatile nature of market returns (Brandt, 1999, Sahalia and Brandt, 2001). Moreover, since we are considering portfolios that involve highly non-normal and mechanically correlated option returns, the estimation issue can only get worse. Eventually, this estimation uncertainty translates into bound uncertainty and may significantly affect our inference. To alleviate this estimation issue, I choose to consider simple static option strategy that has the following form

$$R_P = R_f + \alpha_S(R_S - R_f), \tag{2.16}$$

where α_S denotes the fraction of wealth allocated to a generic return R_S. Hence, only the tradeoff between a safe asset and a return is considered[9].

Figure 2.3 plots the bound frontiers given in inequalities (2.11) and (2.12) when power s equals to 2, 0.5, 0, -1, -3 and -8, and Table 2.2 reports the optimal portfolio weights at a riskfree rate of zero per month[10]. Notice that when $s = 2$, the admissible region for the pricing kernel is below the depicted curves, whereas at other powers it is above the depicted curves. I intentionally leave this "inconvenient" feature in the

[9] In doing this, I also ignore the possible utility gain from a combination of the market index and a derivative strategy. For a CRRA investor with a risk aversion coefficient of more than one, Driessen and Maenhout (2005) show the allocation to the market index is rarely significant. This indirectly shows the limited utility gains by considering the market index.

[10] The mean and the standard deviation of the riskfree rate for the short sample is 4bp per month and 40bp, respectively. Therefore, I center it at zero and extend to ± 3 standard deviations away from the center in Figure 2.3.

Table 2.2: **Optimal portfolio weights for benchmark and index strategies.** Panel A shows the optimal portfolio weights for the optimization problem described in Figure 2.3 at a fixed interest rate of zero. Panel B shows the range of the admissible portfolio weights that guarantees a positive portfolio return series. For a return series $\{R_t\}_{t=1}^T$, the range is given by $[\alpha_{min}, \alpha_{max}] = [1/(1 - \max[\{R_t\}_{t=1}^T]), 1/(1 - \min[\{R_t\}_{t=1}^T])]$.

Power	Market	96%-OTM put	ATM straddle
		Panel A	
$s = 2$	-1.954	0.489	0.149
$s = 0.5$	0.941	-0.199	-0.070
$s = 0$	1.839	-0.341	-0.137
$s = -1$	3.450	-0.437	-0.260
$s = -3$	5.428	-0.442	-0.461
$s = -8$	5.667	-0.442	-0.663
		Panel B	
α_{min}	-9.589	-0.450	-0.680
α_{max}	5.369	1.175	1.281

picture to emphasize the flip in the direction of bounds around $s = 1$. Four different types of portfolios are shown on this plot: two types involve the two benchmark derivative strategies and, for comparison purposes, the other two types involve the market index.

Several patterns emerge from Figure 2.3 and Table 2.2. First, strategies involving the put option clearly dominate the other strategies across all powers and a wide range of riskfree rates. At $s = 2$, which is at the HJ bound, the Sharpe ratio essentially determines the strength of restrictions that a given security return puts on the pricing kernel. As a result, in accordance with the rankings of the absolute Sharpe ratio given in Table 1, 96%-OTM put returns implies the sharpest constraint, and is followed by the index and lastly the straddle returns. At other powers, the optimal bound belongs to the spectrum of generalized entropy bounds and hence has a utility-based interpretation as discussed before. In particular, a bound at a given power s corresponds to the transformed optimized utility of a power utility agent with a risk-aversion of $\frac{1}{1-s}$. In a closely related empirical paper, Driessen and Maenhout (2005) study the asset allocation problem of an investor who has access

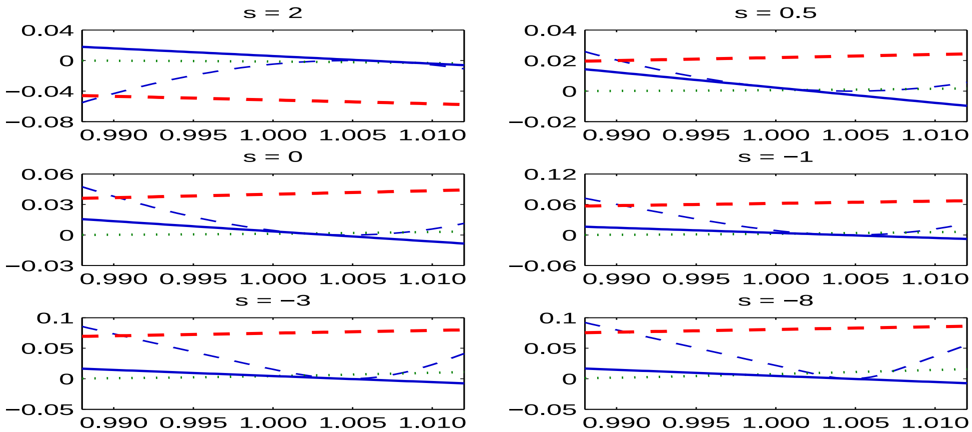


FIGURE 2.3: **Bounds implied by benchmark strategies and the index.** This figure plots the non-parametric bound frontiers (right hand side in inequality (2.11)) for the benchmark trading strategies and the market index across different hypothetical riskfree rates. The estimation is done, at each hypothetical riskfree rate, by conducting a nonlinear search on the optimal portfolio weight α_S to either maximize or minimize the right hand side of inequalities (2.11) and (2.12). The solid line, thin dashed line, dotted line and thick dashed line depict the frontiers for the passive market strategy, active market strategy, ATM straddle strategy and 96%-OTM put option strategy, respectively. The passive market strategy simple sets α_S at zero at every interest rate level and the active market strategy involves a search as described above.

to index options. Their empirical results lend support to my results, especially at $s = 0.5$ and $s = 0$. At $s = 0.5$, both their paper and my results show that an agent with a risk-aversion of two will have a significant short position in the 96%-OTM put option: their paper, allowing the market index to enter into the asset menu as well, has an estimate of about -10% for α_S while I have an estimate of approximately -20% by excluding the index. At $s = 0$, which corresponds to the logrithmic utility case as in the original entropy bound, their estimate is around -15% and mine again roughly doubles their estimate. Putting the difference in asset menus and sample periods aside, both studies show the economic benefits by allowing investors to trade deep OTM put options. The two plots for $s = 0.5$ and $s = 0$ in Figure 2.3 illustrate these benefits by highlighting the differences in utility gains among the four candidate strategies. As s goes negative, the related risk-aversion coefficient becomes even smaller so the relative gain in expected returns by shorting put options further outweighs the loss from variance and other higher-order moments. Consequently, the optimal bound requires even larger absolute position in the OTM put option. Notably, the ATM straddle returns for my sample yield an unimpressive mean relative to 96%-OTM put options and a very high variance relative to the market index. Accordingly, strategies involving the straddle returns appear to imply an inferior bound compared to either the put option strategy or the index strategy.

To offer a deeper interpretation of the above empirical findings, it is worthwhile to repeat the insights in bound interpretations given in Section 2.2. Although the marginal or representative investor determines the market prices of jump and volatility risks[11], investors with different risk-attitudes all reveal valuable information about these prices through their trading behavior. For example, all else being equal, higher prices of jump risks imply more expensive deep OTM put options. With a fixed physical jump distribution, this implies a lower ex-ante and ex-post average return for buying puts. However, for a less risk-averse agent who does not value the put option's hedging ability as much as the average consumer, she treats the increase in put prices as a lucrative trading opportunity. By shorting more, she increases her expected utility. Following this logic, the above empirical findings highlight the more important role of priced jump risks than volatility risks in option prices. Notice that this is not saying that volatility risks should have little pricing impacts, as we are only looking at this through static power utility agents' optimization problems. To have priced volatility risks emerge as an effective bound, we may have to investigate more complicated dynamic strategies given the strong predictability in volatilities. This is left to future research.

Besides the two benchmark derivative strategies, alternative strategies may ap-

[11]The presence of jump and stochastic volatility in the index price is well established in the option pricing literature. Buraschi and Jackwerth (2001), Coval and Shumway (2001) and Bakshi and Kapadia (2003) show the presence of volatility premium. Bates (2002), Pan (2002) and Ait-Sahalia, Wang and Yared (2001) show the presence of jump risk premium.

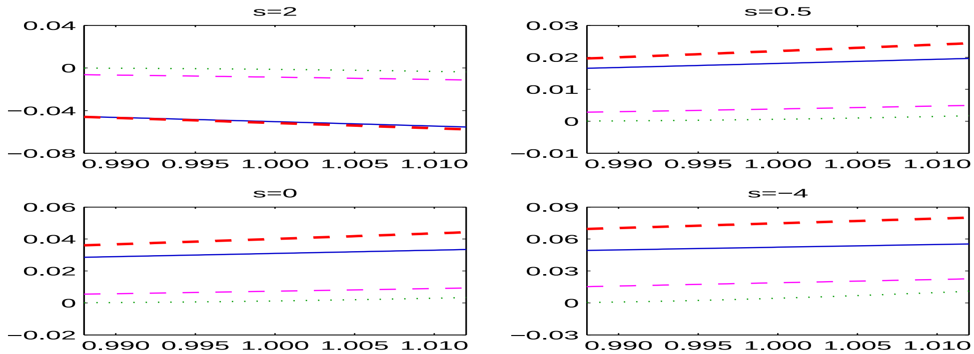


FIGURE 2.4: **Bounds implied by benchmark strategies and two alternative OTM put strategies.** This figure plots the non-parametric bound frontiers (right hand side in inequality (2.11)) for the benchmark trading strategies and two alternative OTM put option strategies across different hypothetical riskfree rates. The estimation is done, at each hypothetical riskfree rate, by conducting a nonlinear search on the optimal portfolio weight α_S to either maximize or minimize the right hand side of inequalities (2.11) and (2.12). The solid line, thin dashed line, dotted line and thick dashed line depict the frontiers for the 92%-OTM put option strategy, ATM put option strategy, ATM straddle strategy and 96%-OTM put option strategy, respectively.

Table 2.3: **Optimal portfolio weights for alternative put and crash-neutral strategies.** Panel A shows the optimal portfolio weights for the optimization problem described in Figure 2.4 and 2.5 at a fixed interest rate of zero. Strategies involving the 92%-put option, ATM put option, crash-neutral put option and crash-neutral straddle are shown. Panel B shows the range of the admissible portfolio weights that guarantees a positive portfolio return series. For a return series $\{R_t\}_{t=1}^T$, the range is given by $[\alpha_{min}, \alpha_{max}] = [1/(1-\max[\{R_t\}_{t=1}^T]), 1/(1-\min[\{R_t\}_{t=1}^T])]$.

Power	92%-OTM put	ATM put	R-C-neutral put	R-C-neutral straddle
		Panel A		
$s = 2$	0.406	0.207	0.253	0.138
$s = 0.5$	-0.129	-0.088	-0.184	-0.070
$s = 0$	-0.202	-0.162	-0.331	-0.138
$s = -4$	-0.253	-0.306	-0.467	-0.545
		Panel B		
α_{min}	-0.254	-0.307	-0.468	-0.601
α_{max}	1.141	1.178	0.253	0.354

pear attractive for certain CRRA investors and thus provide tighter bounds on the pricing kernel. Figure 2.4 and 2.5 show bounds implied by two alternative put option strategies and two crash-neutral strategies, respectively, and Table 2.3 shows the corresponding weights. Figure 2.4 shows that a strategy that shorts the 96%-OTM put option turns out to be the dominating one across all powers. This is somewhat surprising since we would expect its performance to lie between the ones involving

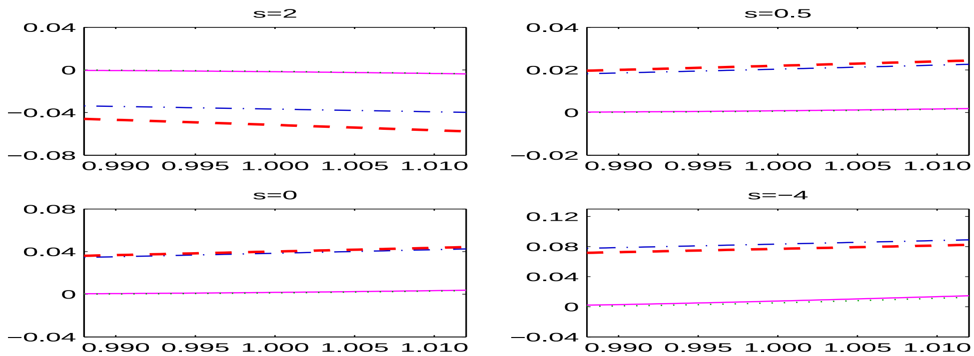


FIGURE 2.5: **Bounds implied by benchmark strategies and two crash-neutral strategies.** This figure plots the non-parametric bound frontiers (right hand side in inequality (2.11)) for the benchmark trading strategies and two crash-neutral strategies across different hypothetical riskfree rates. The estimation is done, at each hypothetical riskfree rate, by conducting a nonlinear search on the optimal portfolio weight α_S to either maximize or minimize the right hand side of inequalities (2.11) and (2.12). Note that the two robust crash-neutral return series are used instead of the original full-sample series. The solid line, dot-dashed line, dotted line and thick dashed line depict the frontiers for the crash-neutral straddle strategy, crash-neutral put option strategy, ATM straddle strategy and 96%-OTM put option strategy, respectively.

the 92%-OTM put and those involving the ATM put. A closer look at the admissible portfolio weights reveal that strategies shorting the 92%-OTM put have strong weight restrictions: given that the maximum net return is close to 4 (See Table 1), the maximum proportion one can short on a 92%-OTM put is $1/(1-(4+1)) = -0.25$ at a riskfree rate of zero. In absolute value, this is significantly smaller than the allowable 0.45 short position on the 96%-OTM put, as shown in Table 2.2. Consequently, despite the fact that the 92%-OTM put has a more negative average return than the 96%-OTM put, weight constraints prevent investors from exploiting it any further. This phenomenon, although statistically irrelevant for the portfolio dominance results (strategies involving the 96%-OTM put are more attractive partly because they allow more aggressive short positions), points to the instability issue of the in-sample portfolio choice problems[12]. I will shortly come back to this issue for a full discussion.

[12] Notice that the in-sample portfolio choice problems are well-defined both theoretically and numerically within our context. In particular, although sometimes the solutions are close to the boundary (See Table 2.2 when $s = -1, -3$ or -8), the infinitely large marginal utility at the boundary for an CRRA investor will restrict the optimal choice to well within the boundary. However, the boundary-dependence of the optimal solution is exacerbated in our context because the CRRA investor's risk-aversion is sometimes close to zero.

Figure 2.5 shows that bounds from the benchmark put option strategy dominate crash-neutral strategies at $s = 2$ and perform well for the other three powers. In particular, the deep OTM put strategy weakly dominates the crash-neutral put strategy at $s = 0.5$ and is on par with the latter at $s = 0$ and a little less informative at $s = -4$. The key observation is that the difference between bounds based on these two strategies is significantly smaller than the difference between them and the other two strategies. Taken as a whole, Figure 2.4 and 2.5 suggest the superior role of deep OTM put strategies (especially strategies with 96%-OTM put) in effectively shaping the admissible space of candidate pricing kernels. They reveal information on the pricing of jump risks in the economy and empirically provide the sharpest bounds that any valid discount rate must satisfy.

The above optimized bounds can be directly used to confront candidate pricing kernels. However, they may appear too stringent or cumbersome from several practical concerns. First, in-sample portfolio choice generates portfolio weights that are too volatile (See Brandt, 2000, Driessen and Maenhout, 2005). What is even more detrimental in our setup is the boundary-dependence of the optimal weights. As seen from Table 2.2 and 2.3, weights for several strategies are close to their boundary values for large negative powers. This exacerbates the in-sample instability issue since extreme observations are more sample-dependent than sample moments. Second, transaction costs and margin requirements for real-world option trading strategies may limit the amount that we can short. Although traction costs are small for the index option market (Bakshi, Cao and Chen, 1997) and our long positions in the riskfree asset can serve as margin, these microstructure effects may become non-negligible if we have excessive short positions on index options. Third, weights for these optimal bounds depend on the prevailing riskfree rate and are thus time-varying. This is cumbersome for most applications, since a different bound has to calculated for a different riskfree rate. Some inspections of Figure 2.4 and 2.5 reveal that option implied bounds are constant across a wide range of bond rates. This prompts us to think about option strategies that are independent of the bond rates.

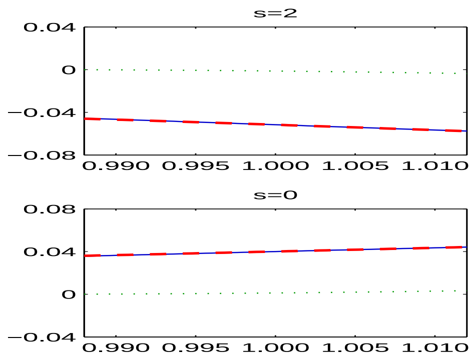


FIGURE 2.6: **Bounds implied by optimal and conservative benchmark strategies.** This figure plots the non-parametric bound frontiers (right hand side in inequality (2.11)) for the benchmark trading strategies and conservative benchmark strategies across different hypothetical riskfree rates. For the optimal benchmark strategies, the estimation is done, at each hypothetical riskfree rate, by conducting a nonlinear search on the optimal portfolio weight α_S to either maximize or minimize the right hand side of inequalities (2.11) and (2.12). The dotted line and the thick dashed line depict the frontier for the ATM straddle strategy and 96%-OTM put option strategy, respectively. The solid lines at $s = 2, 0.5, 0, -4$ depict the interest-rate independent longing 50%, shorting -20%, shorting -35% and shorting -35% on the 96%-OTM put option strategies, respectively.

Driven by the these concerns, I propose a simple way to create conservative and interest rate independent portfolio strategies. For the most informative 96%-OTM put strategies, I set a lower threshold $\alpha_L = -35\%$[13] on the short position to avoid excessive shorting and fix the put weight at the optimal zero-interest put weight given in Table 2.2 if it does not exceed the threshold, or at α_L otherwise. By doing this, we end up having the following three types of OTM put strategies: a long 50% strategy at $s = 2$; a short -20% strategy at $s = 0.5$ and a short -35% strategy at other powers. Figure 2.6 display the bounds for these conservative strategies together with the bounds from the two benchmark strategies. Not surprisingly, they agree well with the optimal strategies for $s = 2, 0.5$ and 0. At $s = -4$, the discrepancy is

[13]This number is chosen to approximately equate the optimal weight at $s = 0$ and a riskfree rate of zero. It is also about 10% away from the boundary weight, which helps alleviate the boundary-dependent problem.

about 1% and much smaller than the difference of around 8% between the straddle strategy and the optimal OTM put strategy. For the rest of the paper, I use these simple yet efficient bounds to study restrictions on the pricing kernel. Again, had we missed important information by utilizing these sub-optimal bounds, they still provide valid, albeit conservative restrictions on the pricing kernel.

2.3.3 Rare disaster models and option return bounds

I consider a representative-agent exchange economy model with infrequent large declines in consumption growth (See Barro, 2006, 2009). More precisely, I focus on models with an iid environment. This is a first step in understanding the distribution of tail events in consumption growth. Moreover, since we restrict ourselves to simple static portfolio strategies in the construction of bounds, it is only fair to consider a pricing kernel with iid shocks. I also adopt a Possion-normal distribution for the jump component in consumption growth[14]. This parametric setup has two appealing features: 1. It is flexible enough to match some of the empirical regularities on rare event distributions as documented by Barro (2006); 2. It is infinitely divisible and thus allows us to "zoom" in on an arbitrarily small frequency. I state the model at the annual frequency but will use its monthly counterpart to match asset market bounds constructed from monthly returns data.

The time-additive utility representation for the representative agent is given by

$$E_0(\sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma}}{1-\gamma}),$$

where γ governs investor risk-aversion. The pricing kernel is known to be

$$\log M_{t+1} = \log \beta - \gamma \log G_{t+1}, \tag{2.17}$$

[14]For its applications in the macro-finance literature, see Naik and Lee (1990), Martin (2007) and Backus, Chernov and Zin (2011)

where $G_{t+1} \equiv C_{t+1}/C_t$ is the consumption growth rate. Log consumption growth is assumed to be driven by two independent shocks,

$$\log G_{t+1} = \epsilon_{t+1} + \eta_{t+1}, \tag{2.18}$$

where $\epsilon_{t+1} \sim \mathcal{N}(\mu, \sigma^2)$ is the normally-distributed component and the distribution of the jump component η_{t+1} is given by

$$\eta_{t+1}|(J = j) \sim \mathcal{N}(j\theta, j\nu^2), J \sim \text{Poisson}(\omega). \tag{2.19}$$

To derive the entropy-related quantities for this kernel, we start from calculating the moment-generating functions (MGF) of the two shocks. The MGF for the normal shock is $E(e^{s\epsilon_{t+1}}) = \exp(\mu s + \sigma^2 s^2/2)$, and the MGF for the Poisson-normal part, shown in Backus, Chernov and Martin (2011), is

$$E(e^{s\eta_{t+1}}) = \exp(\omega[e^{s\theta+(s\nu)^2/2} - 1]). \tag{2.20}$$

Then, by independence of the two shocks, the cumulant-generating function (CGF) can be shown to be

$$CGF(s) \equiv \log E(e^{s \log M_{t+1}}) = s \log \beta - \gamma\mu s + \frac{1}{2}\gamma^2\sigma^2 s^2 + \omega[e^{-\gamma s\theta+(\gamma s\nu)^2/2} - 1]. \tag{2.21}$$

By setting $s = 1$, the continuously compounded one-period riskfree rate $r_f \equiv -\log E(M_{t+1})$ is given by

$$r_f = -(\log \beta - \gamma\mu + \frac{1}{2}\gamma^2\sigma^2 + \omega[e^{-\gamma\theta+(\gamma\nu)^2/2} - 1]). \tag{2.22}$$

Finally, combining the above two pieces, the generalized entropy function (GEF) can be shown as

$$GEF(s) = \frac{1}{2}\gamma^2\sigma^2(1-s) + \omega[e^{-\gamma\theta+(\gamma\nu)^2/2} - \frac{1}{s}e^{-\gamma s\theta+(\gamma s\nu)^2/2} - \frac{s-1}{s}]. \tag{2.23}$$

Table 2.4: **Parameter specifications for disaster models** This table shows the parameter specifications for disaster models. Panel A shows the fixed parameters. The total variance in consumption growth is given by $\sigma^2 + \omega(\theta^2 + \nu^2)$. Panel B shows the disaster intensity and size combinations that represent three types of disaster distributions: light disaster type (ω_L, θ_L), mild disaster type (ω_M, θ_M) and severe disaster type (ω_S, θ_S).

Parameter	Value
Panel A	
β	0.99
γ	5
r_f	0.02
$\sigma^2 + \omega(\theta^2 + \nu^2)$	0.035^2
ν^2	0.2^2
Panel B	
ω_L	0.04
θ_L	-0.15
ω_M	0.02
θ_M	-0.30
ω_S	0.01
θ_S	-0.60

When $s \to 0$, $GEF(s)$ reduces to the original entropy

$$L(M) = \frac{1}{2}\gamma^2\sigma^2 + \omega[e^{-\gamma\theta+(\gamma\nu)^2/2} + \gamma\theta - 1]. \tag{2.24}$$

Using $GEF(s)$ as a model diagnosing tool, we are now ready to explore the effects of rare disasters on the pricing kernel. The stacked parameter vector for a disaster model is $(\beta, \gamma, \mu, \sigma^2, \omega, \theta, \nu^2)'$, which has seven components. To better concentrate on economically interesting parameters such as the disaster intensity ω and size θ, I perform a "partial derivative" exercise. First, as shown in Panel A of Table 2.4, I fix the two preference parameters, the riskfree rate and two variance statistics related to the consumption growth process. The total variance in consumption growth $\sigma^2+\omega(\theta^2+\nu^2)$ is fixed at the sample estimate based on the US real consumption data (Backus, Chernov and Martin (2011)) and the variance for the normal component of the Poisson-normal shock is estimated from realized disasters based on international data (Barro ,2006). These second-order moments can be estimated with much more precision than first-order moments (μ and θ) and my choices agree with the disaster literature. Next, I fix the expected loss in a disaster state at $\omega\theta = -0.006$ and consider three ω and θ combinations that represent light, mild and severe disaster types, respectively, as shown in Panel B of Table.2.4[15] These three types of disaster distributions imply increasing magnitude in disaster size and roughly agree with the historical consumption data of the US, an average country in Barro's sample (Barro, 2006), and a few European countries which experienced large drops in per capita GDP during World War II, respectively. Note that these alternative tail specifications are difficult to differentiate empirically, as by design they imply exactly the same mean consumption growth, total volatility in growth and interest rate. This relates to the so-called pseudo problem in disaster models. It is therefore interesting to see whether GEF can better distinguish these models, and additionally, whether asset market returns provide support for any of them.

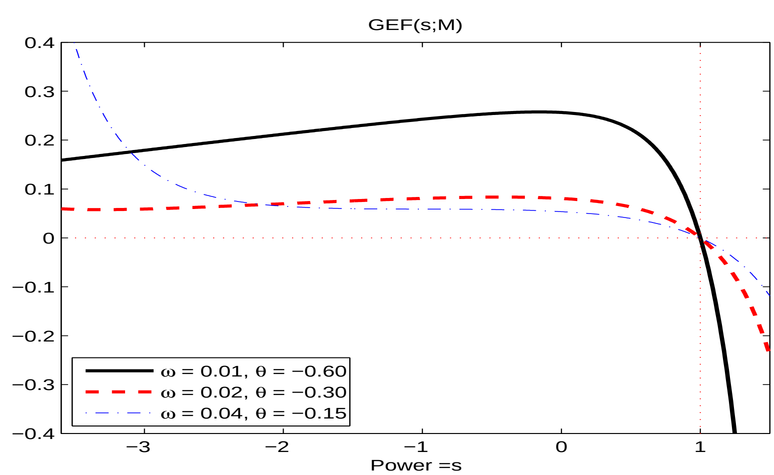


FIGURE 2.7: **Generalized entropy function plots for three disaster models.**

[15]The mean μ of the normal shock component is used to match the interest rate.

Figure 2.7 shows the corresponding GEF plots for the three disaster models. I focus on the region from -3.5 to 1. As s goes beyond 1 or below -3.5, either the severe disaster case or the light disaster case will display explosive behavior. In addition, I will show later that bounds at these powers are less informative anyway. Starting at $s = 1$ where all three GEF's equal zero and going left, the GEF of the severe disaster type rises more steeply compared to the other two and reaches its peak around $s = 0.8$. At its peak, the GEF more than triples that of either the mild or light disaster type. Going further left, it remains the dominating one until s reaches -3, at which the GEF from for light disaster type catches up. The mild disaster type follows a similar pattern, rising more than the light disaster case initially and meeting it at around -2. Eventually, all three start rising sharply for large negative powers, with the light disaster case being the earliest to rise.

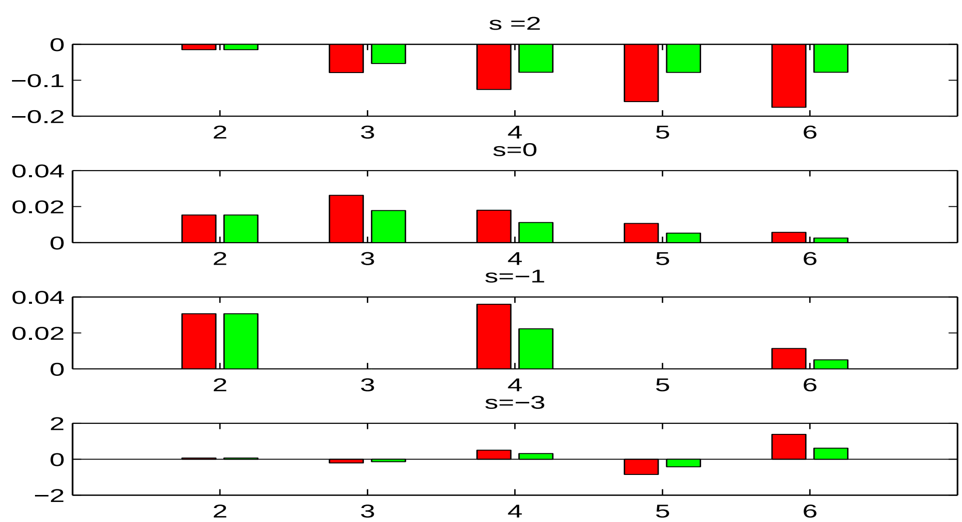


FIGURE 2.8: **Weighted cumulants for two disaster models.** This figure displays the second to sixth weighted cumulants for the mild and light disaster model at $s = 2, 0, -1$ and -3. The j-th weighted cumulant is defined as $\frac{\kappa_j(\log M_{t+1})}{j!}(1-s^{j-1})$ in equation (2.13). The left (dark) bar and the right (light) bar measure the weighted cumulant for the mild and light disaster model, respectively.

To see how various weighting schemes on cumulants generate the patterns in the

GEF, Table 2.8 displays the contributions from the second to sixth weighted cumulants to the overall (generalized) entropy.[16] Since the severe disaster type involves explosive behavior at the sixth moment, I choose to focus on the first two types of disaster distributions. We can view the case at $s = 0$ as the benchmark, since all individual cumulants are weighted equally. At $s = 0$, both disaster types imply the same second cumulant (variance) by design. However, the mild disaster case implies higher third to sixth cumulants. As a result, its overall entropy is higher.[17] At $s = 2$, which corresponds to the Hansen-Jaganathan bound, the signs for cumulants are reversed and, more importantly, higher-order cumulants are magnified compared to the case at $s = 0$. This is because the polynomial coefficients attach more weights to higher-order terms at $s = 2$. This also explains the relatively large magnitude in entropy when the power goes above one, as shown in Figure 2.7. At $s = -1$, interestingly, all odd cumulants vanish and the entropy is a sum of even moments only. Given the vast literature on skewness preferences[18], it is interesting to see whether even moments alone can stand some of the empirical regularities from option returns. In other words, stronger restrictions on the pricing kernel might be found by focusing on $s = -1$ since skewness or in general odd cumulants are no longer present to help boost up entropy. More dramatically, at $s = -3$, odd cumulants show up as negative and hence cancel out the effects of even cumulants. The empirical inquiry at $s = -1$ applies to this case as well. Moreover, the bites coming from negative odd cumulants explain why the light disaster GEF dominates the mild disaster GEF at $s = -3$: rel-

[16]Strictly speaking, as in Equation 2.13, the j-th individual cumulant of the log pricing kernel is $\kappa_j(\log M)$. However, with a slight abuse of terminology, I sometimes refer to the factorial adjusted cumulant $\frac{\kappa_j(\log M)}{j!}$ as cumulant since the factorial provides a natural scaling of the raw cumulant. The j-th weighted cumulant is defined as $\frac{\kappa_2(\log M)}{j!}(1 - s^{j-1})$. See Backus, Chernov and Martin (2011) for the derivation of the individual cumulants for Poisson-normal shocks

[17]Of course, other higher-order cumulants matter, so we need to extrapolate from the patterns seen from the second to sixth cumulant.

[18]See Kraus and Litzenberger (1976), Rubinstein (1973), Harvey and Siddique (2000).

atively larger disaster size θ for the mild disaster case generates disproportionately high absolute odd cumulants that in turn reduce the overall entropy. In sum, the generalized entropy function generates interesting combinations of cumulants and it remains an empirical question whether asset returns can bind it in a meaningful way.

I now confront the rare disaster models with option implied bounds. Analogous to the calibration approach in the macro-finance literature, the plan is to mark up the admissible parameter space corresponding to a set of asset market bounds. Similar approaches have been taken by Hansen and Jaganathan (1991) to depict the efficient mean-variance frontier based on market Sharpe ratios and by Bansal and Lehmann (1994) to restrict the representative agent's risk-aversion based on equity premium. Again, to sharpen our focus on economically interesting quantities, I choose to consider the triple $(\omega, \theta, \gamma)'$. To avoid negative volatility, this time I choose to fix the variance σ^2 of the normal shock component.[19] Also, risk aversion is released as a free parameter to match return bounds. Other than these changes, the rest are the same as in Panel A of Table 2.4.

Bounds based on option strategies essentially delineate a domain in the three-dimensional (ω, θ, γ) space. To ease visual inspection, I plot the contours on the two-dimensional (ω, γ) plane. Figure 2.9 shows these contours for four values of disaster size θ. Many interesting patterns emerge from this figure. First, as expected, larger magnitude of θ requires smaller risk aversion. At $\theta = -0.10$, a 4% annual equity risk premium asks for a risk aversion of around 5.5 if a disaster occurs every 60 years. When θ drops to -0.50, a risk aversion of 2.3 would suffice. Fixing the dis-

[19]We could do this for the previous calibration exercise. In fact, there will be little change in model implications if we adopt this, because only local variants of the baseline disaster model ($|\omega\theta|$ is small) are considered. I choose to fix the total variance before so it has a partial derivative flavor. This time, however, since we are searching over the entire (ω, θ, γ) space, a fixed total variance may sometimes imply a negative σ^2. Therefore, I set σ^2 at 0.035^2 instead. In fact, given the rare occurrence of disasters (especially for the US), I think it makes more sense to equalize σ^2 with the estimate based on historical data.

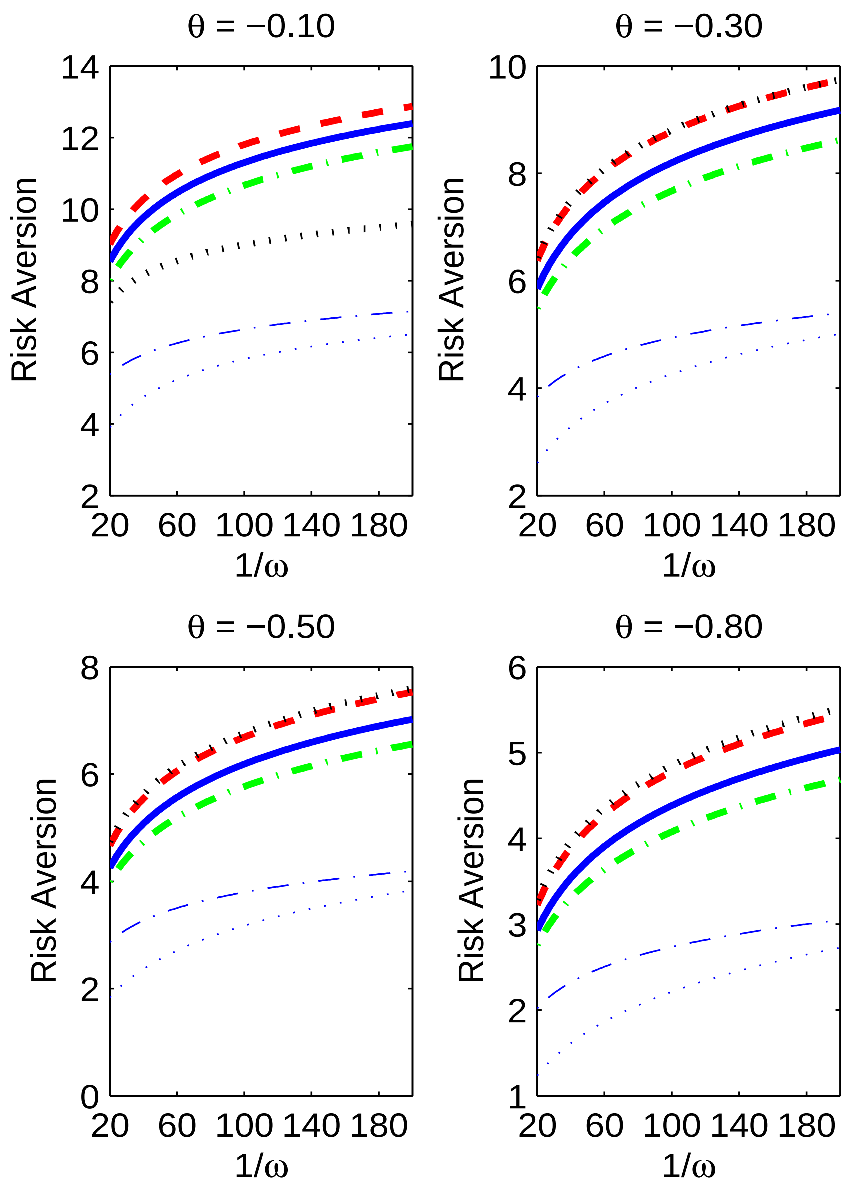


FIGURE 2.9: **Risk aversion bounds implied by index option returns.** This figure shows the required risk aversion coefficients corresponding to different disaster frequency ω, disaster size θ and entropy bounds based on option returns. The thin dotted line depicts the required risk aversion in generating a 4% annual equity risk premium. The thin dash-dotted line, thick dash-dotted line, thick solid line, thick dashed line and thick dotted line depict the required risk aversion coefficients in satisfying the entropy bounds at power $s = 2, 0.5, 0, -1$ and -2, respectively.

aster frequency at 1/60, as θ changes from -0.10 to -0.50, the required risk aversion by the most stringent entropy bound drops from slightly more than 10 to around 6. In particular, at Barro's calibration ($\omega = 1/60, \theta = -0.38$), I calculate that a risk aversion of 7.2 is needed to satisfy the entropy bound at $s = -1$. Such a risk aversion may be regarded as too high to reconcile with many aspects of an individual's risk taking behavior. Secondly, option strategy returns require much larger risk aversion coefficient than the equity risk premium. In particular, at $1/\omega = 100$, the HJ bound implied by longing 50% in the 96%-OTM put typically requires an extra 0.5 units in risk aversion across different disaster size specifications. On top of that, the entropy bound at $s = 0$ (by shorting 40% on 96%-OTM put) asks for an additional 2 units in risk aversion. Finally, the incremental requirement imposed by the most stringent entropy bound at $s = -1$ is around 0.5 to 1 in units of risk aversion. Changes in risk aversion may be hard to quantify economically; a better way to read the economic significance off the figure is to reverse the roles of the x- and y-axis. For instance, when $\theta = -0.50$ and fixing the risk aversion at 6, an entropy bound at $s = -1$ implies that disasters need to happen on average at least once every 50 years, whereas a period of around 100 years would suffice for the entropy bound at $s = 0$. A 50% drop in consumption that happens twice every century is certainly worse than the case with only one drop every century. The difference in bounds' implications is hence economically significant. Thirdly, in unreported results, I consider alternative interest rates and entropy bounds at even more negative powers. An annual riskfree rate in the range of $(1.00, 1.06)$ results in little change in the contour plots. This is mainly because of the insensitivity of option implied bounds to interest rates, as we discussed before. Entropy bounds at more negative powers do not substantially improve on the parameter frontiers depicted by the entropy bound at $s = -1$.

To relate my findings to the literature, it is important to emphasize the methodological differences. In particular, Backus, Chernov and Martin (2011) try to infer rare event information from option data. However, two strong assumptions are made in their paper to make the task feasible: 1. Dividend is a levered claim on consumption; 2. A Merton type option pricing model is chosen to summarize the cross-section of option data. Both are based more on convenience than reality, and it is difficult to measure how small deviations from them could affect the inference on tail information. My approach, on the other hand, is model-free by nature. Based on the fundamental no-arbitrage condition, it asks how much (generalized) dispersion a pricing kernel has to generate in order to rationalize the profits from trading jump risks. Although it cannot deliver a definitive point estimate, informative bounds can tell a lot about the robust features of a pricing kernel. In terms of the empirical findings, they conclude that option prices imply lower probabilities of extreme adverse events than what international macroeconomic data imply. My results, to the contrary, show that more frequent and/or severe disasters are probably needed to reconcile with the historical option return series.[20]

2.3.4 Testing rare disaster models with option market bounds

In this section, I study the statistical significance of the violation of bounds. I start by laying down a testing framework. Suppose the set of parameters governing a specific model is Π. In the case of a disaster model, $\Pi = (\beta, \gamma, \mu, \sigma^2, \omega, \theta, \nu^2)'$. The transformed moment vector of the pricing kernel is defined as

$$\Omega_M(\Pi; S) = \begin{bmatrix} [EM^{s_1}]^{\frac{1}{1-s_1}} \cdot I_{s_1 \in [0,1)} \\ [EM^{s_2}]^{\frac{1}{1-s_2}} \cdot I_{s_2 \in [0,1)} \\ \cdots \\ [EM^{s_N}]^{\frac{1}{1-s_N}} \cdot I_{s_N \in [0,1)} \end{bmatrix}, \tag{2.25}$$

where $S = (s_1, s_2, \ldots, s_N)'$ denotes the collection of powers we are interested in and $I_{s_j \in [0,1)}$ equals -1 if $s_j \in [0, 1)$ and 1 otherwise. At $s = 0$, $I_{s \in [0,1)} = -1$ and the

[20]My results partially agree with the option return literature (See Jackwerth (2000) and Bondarenko (2003)), which find that more frequent crashes need to be added to explain put returns.

corresponding transformed moment $E(M^s)^{\frac{1}{1-s}}$ should be understood as $E(\log M)$. These sign functions, $I_{s_j \in [0,1)}$'s, adjust the directions of bounds so that the left hand side (moments of the pricing kernel) always dominate the right hand side (moments of market returns). Similarly, the transformed return moment vector is given by

$$\Omega_R(R; S) = \begin{bmatrix} E(R_1^{\frac{s_1}{s_1-1}}) \cdot I_{s_1 \in [0,1)} \\ E(R_2^{\frac{s_1}{s_1-1}}) \cdot I_{s_2 \in [0,1)} \\ \cdots \\ E(R_N^{\frac{s_1}{s_1-1}}) \cdot I_{s_N \in [0,1)} \end{bmatrix}, \tag{2.26}$$

where R_j denotes a specific type of return from the market. At $s = 0$, the return moment $E(R^{\frac{s}{s-1}})$ should be treated as $-E(\log R)$. For a valid parameterization Π of the pricing kernel, the difference between $\Omega_M(\Pi; S)$ and $\Omega_R(R; S)$ should be positive, i.e.,

$$\Theta(S) \equiv \Omega_M(\Pi; S) - \Omega_R(R; S) > 0. \tag{2.27}$$

The element-wise positivity of $\Theta(S) = (\theta(s_1), \theta(s_2), \ldots, \theta(s_N))'$ constitutes our basic testable assumptions. In actual estimation, the population moments of returns can be replaced by their sample counterparts for a given sample size T:

$$\hat{\Theta}(S) = \Omega_M(\Pi; S) - \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} R_{t,1}^{\frac{s_1}{s_1-1}} \cdot I_{s_1 \in [0,1)} \\ \sum_{t=}^{T} \frac{1}{T} R_{t,2}^{\frac{s_2}{s_2-1}} \cdot I_{s_2 \in [0,1)} \\ \cdots \\ \sum_{t=1}^{T} \frac{1}{T} R_{t,N}^{\frac{s_N}{s_N-1}} \cdot I_{s_N \in [0,1)} \end{bmatrix}. \tag{2.28}$$

By applying the Generalized Method of Moments of Hansen and Singleton (1982), we can easily find the asymptotic distribution of $\hat{\Theta}(S)$ for this just identified system. However, our problem is a non-standard multivariate inequality test and the usual Wald or Likelihood-ratio tests will not apply. The difficulty lies in the specification of a null hypothesis that can generate easy-to-calculate critical values. Following the multivariate testing literature by Gourieroux, Holly and Monfort (1982), Wolak

(1987) and especially a recent paper by Patton and Timmermann (2010), I set the null at $\Theta(S) = 0$, which is least favorable to the alternative $\Theta(S) > 0$. To find p-values, I simulate a large number of draws from the empirical limiting distribution of $\hat{\Theta}(S)$ and calculate the fraction of draws that result in an element-wise positive θ. This is similar to the simulation exercise in Patton and Timmerman (2010).

The above testing procedure ignores the estimation uncertainty in the GMM asymptotic variance-covariance matrix. This uncertainty could be large given the presence of highly skewed and fat-tailed option returns. As an alternative, I bootstrap the historical return data to provide robust p-values. In particular, I re-sample the historical return series with replacement for a large number of times. For each sample, I calculate the in-sample $\Omega_R(R; S)$ vector and compare it to $\Omega_M(\Pi; S)$. Finally, I count the number of times that $\Omega_M(\Pi; S)$ does not lie above $\Omega_R(R; S)$ in an element-wise sense.

Ideally, we would like to jointly consider generalized entropy bounds at various powers with multiple assets. However, a few statistical issues restrict the way in which we can form these tests. First, testing moment bounds at different powers using the same asset is problematic. This is because the disturbance terms are perfectly related in a nonlinear way, which violates the basic ergodicity assumption necessary for most asymptotic theories to work. Second, the limiting distribution becomes increasingly unstable as we increase the dimension of the testing statistics. Given a few hundred monthly observations, this puts a practical limit on the size of the panel of bounds. Facing these issues, I choose to consider two types of simple bound tests. One is the univariate entropy bound test using market returns only. This also serves as the benchmark test since many recent papers studying equity risk premium impose such a bound (See Backus, Chernov and Martin, 2011, 2012, Martin, 2008 and Alrevaz and Jermann, 2005). The other type is the joint test of the entropy bound of the index and a generalized entropy bound of an option trading strategy.

As in previous sections, bounds at powers of 2, 0.5, 0, -1 and -2 are considered and the corresponding optimal option trading strategies are given in Section 4.2.

To focus on important quantities and to isolate the effects of the riskfree rate, I use the following procedure to generate a null hypothesis for the parameter vector Π. First, similar to before, I treat (β, σ^2, ν^2) as nuisance parameters and set them at $(0.99, 0.035^2, 0.2^2)$. Second, I set the real annualized riskfree rate at 0.95, 1.00 and 1.05, which roughly correspond to the lower bound, mean and upper bound of a sensible estimate based on the US history. Lastly, given a point of interest for the triple $(\omega, \theta, \gamma)'$, I choose μ to match the required interest rate. By doing this, I give $(\omega, \theta, \gamma)'$ a lot of freedom in meeting bounds from option data and the burden in hitting a target interest rate is transferred onto μ. Of course, a large deviation of μ from the historical consumption growth indicates a failure of the hypothesized Π. I choose to report the model-implied mean consumption growth $\mu + \omega\theta$ instead[21].

Table 2.5 reports the testing results for the baseline disaster model with $\omega = 0.02$ and $\theta = -0.35$. This roughly corresponds to the $\omega = 0.017, \theta = -0.38$ estimate based on the empirical distribution of international disasters by Barro (2006), and close to the baseline parameter choice of the rare disaster literature (See Martin, 2009, Backus, Chernov and Zin, 2011). At $\gamma = 5$, both the entropy bound for the index and the HJ bound for the optimal option strategy are satisfied at the historical mean, as demonstrated by the positivity of the corresponding M_{diff} statistics.[22] Notably, at $s = 0$, the model implied entropy is in excess of the risk premium by at least 5% annually. Hence, not surprisingly, the MKT test has p-values well above

[21] This is because $E_c = \mu + \omega\theta$ is the strict model counterpart to the mean historical consumption growth. For reasonable ω and θ pairs, the difference between E_c and μ is small.

[22] At $s = 0$, the M_{diff} statistic can be directly interpreted as the dispersion in the pricing kernel in excess of the risk premium. At other s values, M_{diff} measures generalized excess dispersion and its sign indicates if a bound is satisfied at the historical mean return moment.

Table 2.5: **Baseline disaster model testing results.** This table reports the testing results for the baseline disaster model with $\omega = 0.02, \theta = -0.35$. R_f is the annual riskfree rate and $E_c = \mu + \omega\theta$ is the implied mean consumption growth. MKT denotes the test of the entropy bound with the market return alone, and MKT+OPT denotes the joint test of the entropy bound for the market return and generalized entropy bound at power s for the corresponding option strategy return. M_{diff} is the mean difference given in equation (2.28). P^a-value and P^b-value are the p-values generated from the theoretical limiting distribution and a bootstrapped procedure, respectively.

			MKT	MKT + OPT				
			$s=0$	$s=2$	$s=0.5$	$s=0$	$s=-1$	$s=-2$
$\gamma=2$	$R_f=0.95$	M_{diff}	−0.081	−1.018	−0.315	−0.500	−0.387	−0.566
	(E_c =-0.025)	P^a-value	0.029	0.004	0.001	0.003	0.000	0.000
		P^b-value	0.031	0.005	0.001	0.005	0.000	0.000
	$R_f=1.00$	M_{diff}	−0.030	−1.120	−0.263	−0.449	−0.361	−0.532
	($E_c=0.001$)	P^a-value	0.252	0.017	0.009	0.025	0.001	0.000
		P^b-value	0.234	0.017	0.011	0.027	0.001	0.000
	$R_f=1.05$	M_{diff}	0.019	−1.218	−0.215	−0.400	−0.337	−0.500
	($E_c=0.025$)	P^a-value	0.666	0.023	0.036	0.063	0.002	0.000
		P^b-value	0.664	0.022	0.041	0.066	0.003	0.001
$\gamma=5$	$R_f=0.95$	M_{diff}	0.057	2.698	−0.213	−0.362	−0.303	−0.444
	($E_c=0.018$)	P^a-value	0.906	0.904	0.046	0.096	0.003	0.002
		P^b-value	0.901	0.908	0.052	0.100	0.006	0.002
	$R_f=1.00$	M_{diff}	0.108	2.628	−0.162	−0.311	−0.278	−0.409
	($E_c=0.028$)	P^a-value	0.995	0.994	0.100	0.139	0.010	0.003
		P^b-value	0.995	0.995	0.104	0.135	0.012	0.004
	$R_f=1.05$	M_{diff}	0.157	2.561	−0.114	−0.262	−0.253	−0.376
	($E_c=0.044$)	P^a-value	1.000	1.000	0.179	0.173	0.017	0.005
		P^b-value	1.000	1.000	0.182	0.175	0.020	0.007
$\gamma=6$	$R_f=0.95$	M_{diff}	0.202	9.103	−0.093	−0.217	−0.223	−0.331
	($E_c=0.039$)	P^a-value	1.000	1.000	0.227	0.225	0.030	0.014
		P^b-value	1.000	1.000	0.225	0.216	0.034	0.016
	$R_f=1.00$	M_{diff}	0.253	9.088	−0.043	−0.165	−0.197	−0.296
	($E_c=0.048$)	P^a-value	1.000	1.000	0.364	0.288	0.046	0.024
		P^b-value	1.000	1.000	0.350	0.269	0.047	0.032
	$R_f=1.05$	M_{diff}	0.302	9.074	0.005	−0.116	−0.173	−0.263
	($E_c=0.056$)	P^a-value	1.000	1.000	0.519	0.336	0.071	0.042
		P^b-value	1.000	1.000	0.506	0.322	0.069	0.045
$\gamma=7$	$R_f=0.95$	M_{diff}	0.485	10.831	0.149	0.067	−0.072	−0.121
	($E_c=0.074$)	P^a-value	1.000	1.000	0.885	0.590	0.275	0.202
		P^b-value	1.000	1.000	0.886	0.587	0.257	0.200
	$R_f=1.00$	M_{diff}	0.536	10.831	0.199	0.118	−0.046	−0.086
	($E_c=0.081$)	P^a-value	1.000	1.000	0.944	0.659	0.348	0.292
		P^b-value	1.000	1.000	0.948	0.644	0.338	0.286
	$R_f=1.05$	M_{diff}	0.585	10.831	0.246	0.167	−0.021	−0.052
	($E_c=0.088$)	P^a-value	1.000	1.000	0.976	0.727	0.435	0.361
		P^b-value	1.000	1.000	0.979	0.715	0.416	0.365

0.90 across all interest rate specifications, suggesting no evidence in the rejection of the entropy bound from the index returns. This is also the case when HJ bound is added, indicating no discriminating power from HJ bound either. When we include more demanding option strategy returns at the original entropy bound ($s = 0$), the p-values drop to 10-18%. The reduction in p-value is impressive, especially considering the close to one p-value when the index return is the only testing asset. Nonetheless, the model survives at conventional significance levels. When s goes to -1 and -2, p-values drop to well below 5%, suggesting a strong rejection of the baseline disaster model across all target interest rates.

To remedy this, one must increase the amount of dispersion in the generalized entropy function. As we learned from the previous section (see Figure 2.9), one good way to achieve this is to increase the marginal investor's risk aversion. Indeed, when γ is raised to 6, the model is borderline accepted at the 5% significance level, and, at $\gamma = 7$, the p-values show no sign of rejection at all. However, as risk aversion gets larger, the implied mean growth rate becomes implausibly high. For instance, the implied mean growth rate E_c is about 5% when $\gamma = 6$ and $R_f = 1.00$; when γ rises to seven, E_c is in the range of 7-9%. These numbers are in contradiction with the historical consumption data for the US. Clearly, a tension exists between risk aversion and the riskfree rate, and it is stronger than what thin-tailed consumption growth models imply. An inspection of equation (2.22) tells us why this is so. Compared to standard models with normal shocks, disaster models carry an extra exponential term $\omega[e^{-\gamma\theta+(\gamma\nu)^2/2} - 1]$. Since disasters rarely happen (ω is small), this term is small for low γ values. However, as risk aversion rises, it grows exponentially and quickly dominates the intertemporal substitution effect $\gamma\mu$ and the precautionary savings effect $\frac{1}{2}\gamma^2\sigma^2$. In particular, at $\gamma = 7$ and assuming a mean consumption growth rate of 2%, this term is four times the value of the substitution effect and 20 times the value of the precautionary savings effect. Clearly, under a disaster model framework, the agent's hedging demand for rare event risk is strong for even moderately high risk aversion levels. Given a mean consumption growth around 2%, this strong hedging motive requires low risk aversion to reconcile with the relatively high interest rate. On the other hand, asset market returns ask for higher risk aversion to meet generalized entropy bounds. These two forces make the determination of the representative agent's risk attitude no longer a one-sided exercise. Taken as a whole, it seems that the baseline disaster model is unable to explain the bond market and the option market at the same time.

Table 2.6: **US type disaster model testing results.** This table reports the testing results for the baseline disaster model with $\omega = 0.02, \theta = -0.10$. R_f is the annual riskfree rate and $E_c = \mu + \omega\theta$ is the implied mean consumption growth. MKT denotes the test of the entropy bound with the market return alone, and MKT+OPT denotes the joint test of the entropy bound for the market return and generalized entropy bound at power s for the corresponding option strategy return. M_{diff} is the mean difference given in equation (2.28). P^a-value and P^b-value are the p-values generated from the theoretical limiting distribution and a bootstrapped procedure, respectively.

			MKT	MKT + OPT				
			$s=0$	$s=2$	$s=0.5$	$s=0$	$s=-1$	$s=-2$
$\gamma=2$	$R_f=0.95$	M_{diff}	−0.089	−1.054	−0.319	−0.507	−0.392	−0.574
	$(E_c=-0.028)$	P^a-value	0.020	0.002	0.001	0.003	0.000	0.000
		P^b-value	0.023	0.004	0.001	0.004	0.000	0.000
	$R_f=1.00$	M_{diff}	−0.037	−1.156	−0.268	−0.456	−0.366	−0.540
	$(E_c=-0.003)$	P^a-value	0.196	0.012	0.008	0.019	0.001	0.000
		P^b-value	0.191	0.013	0.010	0.023	0.001	0.000
	$R_f=1.05$	M_{diff}	0.012	−1.254	−0.219	−0.407	−0.342	−0.507
	$(E_c=0.022)$	P^a-value	0.601	0.016	0.036	0.060	0.001	0.000
		P^b-value	0.610	0.017	0.037	0.062	0.003	0.000
$\gamma=5$	$R_f=0.95$	M_{diff}	−0.054	−0.731	−0.298	−0.472	−0.364	−0.518
	$(E_c=-0.004)$	P^a-value	0.107	0.023	0.003	0.010	0.000	0.000
		P^b-value	0.110	0.027	0.006	0.013	0.001	0.001
	$R_f=1.00$	M_{diff}	−0.003	−0.831	−0.247	−0.421	−0.338	−0.483
	$(E_c=0.006)$	P^a-value	0.468	0.060	0.019	0.047	0.001	0.000
		P^b-value	0.473	0.062	0.020	0.048	0.002	0.001
	$R_f=1.05$	M_{diff}	0.046	−0.926	−0.198	−0.372	−0.314	−0.451
	$(E_c=0.016)$	P^a-value	0.860	0.060	0.052	0.087	0.005	0.002
		P^b-value	0.862	0.064	0.060	0.090	0.005	0.003
$\gamma=7$	$R_f=0.95$	M_{diff}	0.010	2.272	−0.252	−0.409	−0.320	−0.387
	$(E_c=0.006)$	P^a-value	0.595	0.590	0.019	0.054	0.002	0.004
		P^b-value	0.588	0.584	0.024	0.060	0.005	0.005
	$R_f=1.00$	M_{diff}	0.061	2.199	−0.201	−0.357	−0.294	−0.352
	$(E_c=0.013)$	P^a-value	0.925	0.921	0.053	0.094	0.006	0.009
		P^b-value	0.927	0.928	0.058	0.099	0.009	0.012
	$R_f=1.05$	M_{diff}	0.110	2.128	−0.152	−0.309	−0.26	−0.319
	$(E_c=0.020)$	P^a-value	0.995	0.994	0.111	0.135	0.012	0.015
		P^b-value	0.995	0.994	0.116	0.135	0.013	0.019
$\gamma=9$	$R_f=0.95$	M_{diff}	0.167	10.814	−0.123	−0.252	−0.221	0.418
	$(E_c=0.022)$	P^a-value	1.000	1.000	0.162	0.186	0.031	0.998
		P^b-value	1.000	1.000	0.160	0.185	0.032	0.998
	$R_f=1.00$	M_{diff}	0.218	10.813	−0.073	−0.201	−0.195	0.455
	$(E_c=0.028)$	P^a-value	1.000	1.000	0.289	0.239	0.044	0.999
		P^b-value	1.000	1.000	0.274	0.232	0.050	1.000
	$R_f=1.05$	M_{diff}	0.267	10.813	−0.025	−0.152	−0.171	0.490
	$(E_c=0.033)$	P^a-value	1.000	1.000	0.417	0.297	0.070	0.999
		P^b-value	1.000	1.000	0.409	0.286	0.072	1.000

If the extrapolated disaster distributions cannot explain the US market, what can? I next test disaster models with alternative distributional assumptions. Table 2.6 reports the testing results at $\omega = 0.02, \theta = -0.10$, which is arguably what the

Table 2.7: **Severe type disaster model testing results.** This table reports the testing results for the baseline disaster model with $\omega = 0.01, \theta = -0.60$. R_f is the annual riskfree rate and $E_c = \mu + \omega\theta$ is the implied mean consumption growth. MKT denotes the test of the entropy bound with the market return alone, and MKT+OPT denotes the joint test of the entropy bound for the market return and generalized entropy bound at power s for the corresponding option strategy return. M_{diff} is the mean difference given in equation (2.28). P^a-value and P^b-value are the p-values generated from the theoretical limiting distribution and a bootstrapped procedure, respectively.

			MKT	MKT + OPT				
			$s=0$	$s=2$	$s=0.5$	$s=0$	$s=-1$	$s=-2$
$\gamma=2$	$R_f=0.95$	M_{diff}	-0.077	-0.974	-0.312	-0.496	-0.385	-0.563
	$(E_c=-0.023)$	P^a-value	0.037	0.007	0.001	0.005	0.000	0.000
		P^b-value	0.034	0.006	0.001	0.004	0.000	0.000
	$R_f=1.00$	M_{diff}	-0.026	-1.075	-0.260	-0.444	-0.359	-0.529
	$(E_c=0.003)$	P^a-value	0.281	0.018	0.009	0.027	0.000	0.000
		P^b-value	0.278	0.022	0.011	0.034	0.001	0.000
	$R_f=1.05$	M_{diff}	0.023	-1.172	-0.212	-0.396	-0.334	-0.497
	$(E_c=0.028)$	P^a-value	0.703	0.020	0.037	0.067	0.002	0.001
		P^b-value	0.703	0.029	0.046	0.072	0.003	0.001
$\gamma=5$	$R_f=0.95$	M_{diff}	0.213	9.786	-0.076	-0.206	-0.224	-0.339
	$(E_c=0.049)$	P^a-value	1.000	1.000	0.279	0.230	0.028	0.012
		P^b-value	1.000	1.000	0.262	0.232	0.031	0.014
	$R_f=1.00$	M_{diff}	0.264	9.777	-0.026	-0.154	-0.198	-0.305
	$(E_c=0.059)$	P^a-value	1.000	1.000	0.417	0.286	0.049	0.019
		P^b-value	1.000	1.000	0.413	0.291	0.053	0.027
	$R_f=1.05$	M_{diff}	0.313	9.768	0.022	-0.106	-0.174	-0.271
	$(E_c=0.069)$	P^a-value	1.000	1.000	0.566	0.349	0.066	0.033
		P^b-value	1.000	1.000	0.554	0.334	0.075	0.037
$\gamma=5.5$	$R_f=0.95$	M_{diff}	0.379	10.818	0.068	-0.040	-0.137	-0.220
	$(E_c=0.075)$	P^a-value	1.000	1.000	0.707	0.438	0.123	0.072
		P^b-value	1.000	1.000	0.699	0.427	0.125	0.076
	$R_f=1.00$	M_{diff}	0.430	10.817	0.118	0.011	-0.111	-0.185
	$(E_c=0.084)$	P^a-value	1.000	1.000	0.821	0.512	0.172	0.110
		P^b-value	1.000	1.000	0.828	0.499	0.175	0.102
	$R_f=1.05$	M_{diff}	0.479	10.817	0.165	0.060	-0.086	-0.151
	$(E_c=0.093)$	P^a-value	1.000	1.000	0.906	0.587	0.228	0.158
		P^b-value	1.000	1.000	0.904	0.574	0.220	0.156

US has experienced[23]. Table 2.7 and 2.8 show testing results based on perturbations around the baseline case. In Table 2.6, as expected, the rejections from violations of the entropy bounds are generally stronger than those in the baseline case. In fact, risk aversion has to go all the way up to 9 to pass the bound test at $s = -1$ at 5% level. At the same time, the hedging demand at $\theta = -0.10$ is substantially lower than that at $\theta = -0.35$. As a result, the US type disaster model still implies a sensible mean consumption growth even at $\gamma = 9$. If one is willing to accept such a high risk

[23]For the US, a consumption decline in the magnitude of 10% only happened once: in 1931, the per capita consumption dropped by 9.9%. Hence, strictly speaking, my assumption on the disaster frequency doubles what the US actually experienced. A lower assumed ω value makes the rejections in Table 2.6 even more stronger.

Table 2.8: **Mild type disaster model testing results.** This table reports the testing results for the baseline disaster model with $\omega = 0.04, \theta = -0.15$. R_f is the annual riskfree rate and $E_c = \mu + \omega\theta$ is the implied mean consumption growth. MKT denotes the test of the entropy bound with the market return alone, and MKT+OPT denotes the joint test of the entropy bound for the market return and generalized entropy bound at power s for the corresponding option strategy return. M_{diff} is the mean difference given in equation (2.28). P^a-value and P^b-value are the p-values generated from the theoretical limiting distribution and a bootstrapped procedure, respectively.

			MKT	MKT + OPT				
			$s=0$	$s=2$	$s=0.5$	$s=0$	$s=-1$	$s=-2$
$\gamma=2$	$R_f=0.95$	M_{diff}	−0.085	−1.039	−0.317	−0.503	−0.389	−0.569
	$(E_c=-0.026)$	P^a-value	0.025	0.003	0.001	0.003	0.000	0.000
		P^b-value	0.028	0.004	0.002	0.003	0.000	0.000
	$R_f=1.00$	M_{diff}	−0.033	−1.141	−0.266	−0.452	−0.363	−0.535
	$(E_c=-0.001)$	P^a-value	0.222	0.013	0.007	0.022	0.000	0.000
		P^b-value	0.215	0.014	0.010	0.026	0.001	0.000
	$R_f=1.05$	M_{diff}	0.016	−1.239	−0.217	−0.403	−0.339	−0.502
	$(E_c=0.024)$	P^a-value	0.633	0.018	0.036	0.060	0.001	0.000
		P^b-value	0.634	0.021	0.038	0.066	0.003	0.001
$\gamma=5$	$R_f=0.95$	M_{diff}	−0.009	−0.011	−0.266	−0.427	−0.336	−0.477
	$(E_c=0.005)$	P^a-value	0.426	0.249	0.010	0.043	0.001	0.001
		P^b-value	0.415	0.250	0.014	0.045	0.002	0.002
	$R_f=1.00$	M_{diff}	0.043	−0.104	−0.215	−0.376	−0.310	−0.442
	$(E_c=0.015)$	P^a-value	0.841	0.396	0.040	0.080	0.004	0.001
		P^b-value	0.840	0.383	0.048	0.088	0.006	0.003
	$R_f=1.05$	M_{diff}	0.092	−0.193	−0.167	−0.327	−0.286	−0.409
	$(E_c=0.025)$	P^a-value	0.983	0.376	0.091	0.117	0.007	0.002
		P^b-value	0.984	0.361	0.096	0.122	0.008	0.003
$\gamma=7$	$R_f=0.95$	M_{diff}	0.159	7.682	−0.136	−0.260	−0.235	−0.274
	$(E_c=0.027)$	P^a-value	1.000	1.000	0.139	0.185	0.027	0.032
		P^b-value	1.000	1.000	0.138	0.179	0.026	0.039
	$R_f=1.00$	M_{diff}	0.210	7.655	−0.085	−0.208	−0.209	−0.239
	$(E_c=0.035)$	P^a-value	1.000	1.000	0.253	0.231	0.035	0.053
		P^b-value	1.000	1.000	0.244	0.222	0.041	0.056
	$R_f=1.05$	M_{diff}	0.259	7.629	−0.037	−0.159	−0.184	−0.206
	$(E_c=0.042)$	P^a-value	1.000	1.000	0.392	0.296	0.060	0.086
		P^b-value	1.000	1.000	0.376	0.274	0.058	0.088
$\gamma=8$	$R_f=0.95$	M_{diff}	0.335	10.803	0.012	−0.083	−0.135	−0.012
	$(E_c=0.046)$	P^a-value	1.000	1.000	0.533	0.381	0.122	0.459
		P^b-value	1.000	1.000	0.517	0.369	0.125	0.468
	$R_f=1.00$	M_{diff}	0.387	10.803	0.062	−0.032	−0.108	0.024
	$(E_c=0.052)$	P^a-value	1.000	1.000	0.683	0.454	0.178	0.560
		P^b-value	1.000	1.000	0.689	0.443	0.178	0.555
	$R_f=1.05$	M_{diff}	0.435	10.803	0.109	0.017	−0.084	0.058
	$(E_c=0.058)$	P^a-value	1.000	1.000	0.811	0.525	0.241	0.649
		P^b-value	1.000	1.000	0.806	0.508	0.230	0.640

aversion level, then the US type disaster specification can fit the asset markets. For the severe type model in Table 2.7, a risk aversion of 5.5 suffices to satisfy option return bounds, but a 8% mean growth rate in consumption seems implausibly high. Turning to the less severe but more frequent type shown in Table 2.8, a somewhat high mean growth rate around 3.5% and a somewhat high risk aversion of $\gamma = 7$ are simultaneously needed to pass all the tests.

Table 2.9: **Robust option strategies testing results.** This table reports the testing results for various disaster models with robust option strategies. In particular, the short position in 96%-OTM put is halved to 20% for option strategies at $s = 0, -1$ and -2. R_f is the annual riskfree rate and $E_c = \mu + \omega\theta$ is the implied mean consumption growth. MKT denotes the test of the entropy bound with the market return alone, and MKT+OPT denotes the joint test of the entropy bound for the market return and generalized entropy bound at power s for the corresponding option strategy return. M_{diff} is the mean difference given in equation (2.28). P^a-value and P^b-value are the p-values generated from the theoretical limiting distribution and a bootstrapped procedure, respectively.

			MKT	MKT + OPT				
			$s=0$	$s=2$	$s=0.5$	$s=0$	$s=-1$	$s=-2$
$(\omega=0.02, \theta=-0.35)$	$R_f=0.95$	M_{diff}	0.057	2.698	−0.213	−0.286	−0.151	−0.198
	$(E_c=0.018)$	P^a-value	0.901	0.907	0.045	0.006	0.003	0.002
		P^b-value	0.906	0.908	0.045	0.008	0.004	0.003
	$R_f=1.00$	M_{diff}	0.108	2.628	−0.162	−0.234	−0.125	−0.164
	$(E_c=0.028)$	P^a-value	0.993	0.994	0.097	0.019	0.011	0.015
		P^b-value	0.995	0.995	0.102	0.022	0.013	0.013
	$R_f=1.05$	M_{diff}	0.157	2.561	−0.114	−0.186	−0.101	−0.131
	$(E_c=0.038)$	P^a-value	1.000	1.000	0.186	0.047	0.032	0.032
		P^b-value	1.000	1.000	0.180	0.055	0.033	0.040
$\omega=0.02, \theta=-0.10$	$R_f=0.95$	M_{diff}	−0.054	−0.731	−0.298	−0.396	−0.212	−0.272
	$(E_c=-0.004)$	P^a-value	0.106	0.025	0.003	0.000	0.000	0.000
		P^b-value	0.112	0.026	0.004	0.000	0.000	0.000
	$R_f=1.00$	M_{diff}	−0.003	−0.831	−0.247	−0.345	−0.186	−0.238
	$(E_c=0.006)$	P^a-value	0.474	0.061	0.016	0.001	0.000	0.000
		P^b-value	0.478	0.059	0.019	0.002	0.001	0.001
	$R_f=1.05$	M_{diff}	0.046	−0.926	−0.198	−0.296	−0.162	−0.205
	$(E_c=0.016)$	P^a-value	0.863	0.063	0.054	0.004	0.001	0.002
		P^b-value	0.857	0.058	0.062	0.007	0.003	0.004
$\omega=0.04, \theta=-0.15$	$R_f=0.95$	M_{diff}	−0.009	−0.011	−0.266	−0.351	−0.184	−0.232
	$(E_c=0.005)$	P^a-value	0.427	0.256	0.013	0.001	0.000	0.001
		P^b-value	0.419	0.246	0.014	0.001	0.001	0.001
	$R_f=1.00$	M_{diff}	0.043	−0.104	−0.215	−0.300	−0.158	−0.197
	$(E_c=0.015)$	P^a-value	0.842	0.383	0.041	0.004	0.002	0.004
		P^b-value	0.841	0.374	0.049	0.006	0.003	0.003
	$R_f=1.05$	M_{diff}	0.092	−0.193	−0.167	−0.251	−0.133	−0.164
	$(E_c=0.025)$	P^a-value	0.984	0.375	0.097	0.013	0.006	0.009
		P^b-value	0.984	0.365	0.095	0.018	0.008	0.014
$\omega=0.01, \theta=-0.60$	$R_f=0.95$	M_{diff}	0.213	9.786	−0.076	−0.129	−0.072	−0.094
	$(E_c=0.049)$	P^a-value	1.000	1.000	0.273	0.123	0.091	0.095
		P^b-value	1.000	1.000	0.269	0.129	0.094	0.094
	$R_f=1.00$	M_{diff}	0.264	9.777	−0.026	−0.078	−0.046	−0.059
	$(E_c=0.059)$	P^a-value	1.000	1.000	0.418	0.243	0.193	0.209
		P^b-value	1.000	1.000	0.408	0.240	0.204	0.199
	$R_f=1.05$	M_{diff}	0.313	9.768	0.022	−0.029	−0.022	−0.026
	$(E_c=0.069)$	P^a-value	1.000	1.000	0.569	0.402	0.341	0.361
		P^b-value	1.000	1.000	0.560	0.396	0.338	0.347

Finally, one may wonder whether the demanding entropy bounds at negative powers come from the excessive short position that we allow investors to hold. In other words, given the statistical uncertainty around the optimal allocation rule, maybe the selected representative option trading strategies imply in-sample moment characteristics that are too harsh for any representative agent type of model to satisfy. To address this robustness concern, I reduce α_L by half and redo all tests at $\gamma = 5$. Table 2.9 shows the results. We see that although all the p-values associated with $s = 0, -1$ and -2 are somewhat larger than their counterparts in the previous tables, the rejections are still strong. In particular, the baseline model is rejected at below 5% significance level for $s = -1$ and $s = -2$ across all target interest rates. For the US type and the mild type disaster models, none of the specifications can pass the generalized entropy bound tests at $s = 0, -1$ or -2. For the severe type, the p-values are now well above 5% but it still does not qualify as a successful model since the implied mean consumption growth rate is too high.

As we reduce the magnitude of α_L, we do see that the bound mean statistics M_{diff} gets better (closer to zero) at negative powers. For instance, for the baseline model, the bound inequality at $s = 0$ is improved by around 8%-10%. This is equivalent to saying that we reduce the mean (log) return of the representative option trading strategy by 8%-10%. With such a significant drop in mean, how come the p-values do not increase much? This stems from the small variation of the transformed option returns. To see this, note that at negative power s, we are essentially taking fractional powers of the returns, e.g., at $s = -1, E(R^{\frac{s}{s-1}}) = E(R^{\frac{1}{2}})$. Consequently, the highly volatile option strategy returns are flattened out by these power transformations. This reduced variation in the transformed sample, together with a negative mean bound estimate, makes the rejection of bounds highly significant. This feature again highlights the discriminatory power of generalized entropy bounds at negative powers.

To summarize the above empirical findings, I show how standard disaster models under several parameterizations fail to meet the nonparametric bounds based on robust option trading strategies. The discriminatory power of bounds with negative powers are highlighted. However, I consider these findings as suggestive as opposed to conclusive. First, although I find evidence against streamline disaster models, their abilities in magnifying (generalized) entropy through tail distortions are impressive. Within a risk aversion of ten, even the US type specification can meet all the entropy bounds with a reasonable amount of mean consumption growth. This leads one to conjecture that more sophisticated variants of disaster models may be up to the challenge (See Barro and Ursua, 2008 and Watcher, 2008). Second, even for the current version of disaster model, I have not exhausted all plausible parameter choices. For instance, the variance ν^2 for the normally distributed individual jump is shown to have important effects on the riskfree rate at a high risk aversion level. Yet I set it at 0.2 for simplicity. A more extensive exploration of the seven-component parameter vector Π may yield a winner.

Despite these caveats, there are a few important takeaways from the above exercise. First and foremost, confronting a pricing kernel with the equity risk premium alone is not enough, especially when tail behavior is considered. In fact, except at $\gamma = 2$ and for a riskfree rate below 1, the equity risk premium constraint is satisfied across all specifications, most of the time with a p-value close to one. This reveals its lack of power in discriminating tail behaviors of the pricing kernel. By subjecting a discount rate to a spectrum of option trading strategies at different powers, we gain a better sense of its all-around performance. Secondly, it is crucial to consider generalized entropy bounds at negative powers, not only because of its informativeness by mean moment restrictions as demonstrated by Figure 2.9, but also because of the statistical powers they afford through fractional power transformations of returns. This analytical feature, combined with the moment characteristics of option returns, can potentially provide the most exacting and thus informative moment constraints on the pricing kernel. Of course, return robustness should be checked to avoid overly restrictive constraints.

2.4 Conclusion

Under the fundamental no-arbitrage condition, this paper develops a spectrum of new nonparametric bounds that significantly enrich the nonparametric bound universe. These bounds essentially describe the discrepancy between what an optimizing agent could achieve if all admissible assets were tradable and what she actually achieves in the real-world market, thus providing an economically meaningful way to restrict candidate pricing kernels. Motivated by these new bounds, I propose to use the generalized entropy function, a natural extension of the original entropy, to systematically study market implied bounds. Through cumulant-expansions on both sides of the generalized entropy bounds, I show how the new bounds provide unique information about the pricing kernel. Their abilities in teasing out tail information are also highlighted.

Equipped with these analytical tools, I study index option returns, since their unique moment characteristics can potentially provide the sharpest restrictions on the pricing kernel. Empirically, I find that strategies with short positions in deep OTM put options dominate both the market index and other standard derivative trading strategies in constraining moments of the pricing kernel. This highlights the pricing of jump risks in the index option, and is expected to be useful for inferring rare event distributions in the pricing kernel. I then postulate a pricing kernel in the form of a standard disaster model and use option return bounds to differentiate among alternative parameterizations. Both point estimates and formal testing results indicate the deficiency of standard disaster models in reconciling with option data. Both tail distortion and time-dependency might be needed to attain bounds implied by option returns.

A study of the joint behavior of time-varying disaster distribution and option returns is an obvious avenue to pursue. This not only helps achieve unconditional option return bounds but also generates insights into the time series properties of option returns. The newly developed bound system, in particular an extended version that can cope with conditioning information, is expected to be instrumental. On the other hand, it remains interesting to see other applications of the generalized entropy bounds, possibly on welfare analysis, model diagnosis, and the creation of new representative agent models.

3

Diagnosing Dynamic Asset Pricing Models with Generalized Entropy Bounds

3.1 Theory

3.1.1 Entropy in BCZ

To relate properties of asset returns to features of pricing kernels, BCZ relies on an entropy concept that has wide applications in science and information theory. In particular, for a positive random variable M (e.g., a stochastic discount factor), the entropy is defined as

$$EN(M) = \log EM - E \log M. \tag{3.1}$$

The first term $\log EM$ is revealed by the yield on a riskless bond and $-E \log M$, by Jensen's Inequality and under the no-arbitrage condition, is bounded by the maximal $E \log R$ (expected log return) over a portfolio of assets. Taken together, the entropy must be greater than or equal to the expected excess return. This fundamental link between the pricing kernel and asset returns prompts BCZ to use the entropy as an alternative measure of pricing kernel dispersion.

Under log-normality of the pricing kernel, $EN(M)$ reduces to one-half the variance of the log pricing kernel. More generally, BCZ show that $EN(M)$ is a weighted average of high-order cumulants of the pricing kernel. Let κ_j be the j-th cumulant of the pricing kernel,[1] then $EN(M)$ can be decomposed as

$$EN(M) = \sum_{j=2}^{\infty} \kappa_j(\log M)/j!. \tag{3.2}$$

This decomposition shows how high-order cumulants enter into the entropy definition and is important for the characterization of pricing kernels that feature tail event risks, e.g., disaster models.

3.1.2 Generalized entropy

While BCZ's definition of entropy is intuitively appealing, questions remain about the specific functional form and the economic content of their entropy definition. In particular, in the case of log-normality, why would one-half the variance be the most interesting case? Do there exist other scalings of the variance that are both mathematically tractable and economically informative (i.e., allowing us to study the pricing kernel through financial market data)? In general, why is BCZ's way of weighting the cumulants most interesting?

I attempt to answer the above inquiries using the extended entropy concept developed in Liu (2012). Liu (2012) extends the basic entropy by allowing a more general weighting scheme on the cumulants of the pricing kernel. In particular, Liu (2012) defines the generalized entropy function (GEF) as

$$GEF(M; s) = \log E(M) - \frac{1}{s} \log E(M^s), s \in (-\infty, 1). \tag{3.3}$$

It generalizes the basic entropy as Liu (2012) shows that $GEF(M; s)$ converges to $EN(M)$ when $s \to 0$. In general, Liu (2012) shows that $GEF(M; s)$ can be decom-

[1] Cumulants are slightly different from moments. For the exact calculation of cumulants, see Backus, Chernov and Martin (2011).

posed as

$$GEF(M;s) = \sum_{j=2}^{\infty} \frac{\kappa_j(\log M)}{j!}(1 - s^{j-1}). \tag{3.4}$$

Setting s at zero, we arrive at BCZ's entropy definition. In the case of normality, it can be shown that

$$GEF(M;s) = \frac{1}{2}(1 - s)Var(\log M). \tag{3.5}$$

We therefore obtain scalings that are different from one-half as in the basic entropy definition. In general, equation (3.4) permits a polynomial weighting of the individual cumulants. By varying the power s, we gain insights into various combinations of the moments of the pricing kernel. For example, by setting s at large negative numbers, GEF roughly calculates the difference between even and odd moments of the pricing kernel. This difference is important for diagnosing pricing kernels that feature asymmetric jumps (e.g., consumption-based models that feature rare disasters often include downside jumps only). BCZ's definition sets s at zero and is unable to capture this information. Liu (2012) builds on this insight to make inference about a disaster model by using option market returns.

More importantly, Liu (2012) shows that similar to the entropy inequality, the generalized entropy function at power s is bounded from below by moments of asset returns:

$$GEF(M;s) \geqslant \frac{s-1}{s} \log E(R^{\frac{s}{s-1}}) - \log(R_f), s \in (-\infty, 1), \tag{3.6}$$

where R_f is the riskfree rate and R is an arbitrary return. If we define $\gamma = 1/(1 - s)$, then $E(R^{\frac{s}{s-1}})$ equals $E(R^{1-\gamma})$, which is the expected utility (up to a scaling constant) of an agent with a risk aversion coefficient of γ. Therefore, the generalized entropy function at power s ($GEF(M;s)$) is no less than the "excess" utility (i.e., the right hand side of inequality (3.6)) for an agent with risk aversion $\gamma = 1/(1 - s)$. This interpretation parallels the interpretation of the basic entropy inequality and allows us to directly link pricing kernel dispersion (i.e., generalized entropy) to return moments. For more properties on generalized entropy, see Liu (2012).

3.1.3 Diagnosing Asset Pricing Models Using Generalized Entropies

What does generalized entropy give us in diagnosing asset pricing models? Or more specifically, what information about the pricing kernel is revealed through generalized entropy that is not already in entropy? In this section, I discuss the power of generalized entropy, especially in comparison with the basic entropy, from a theoretical point of view.

Time-homogeneous Discount Factor

To set up the framework for a time-homogeneous stochastic discount factor, we consider a general representation of the pricing kernel that includes many existing consumption-based asset pricing models as special cases. Suppose the pricing kernel is given by

$$M_{t+1} = \exp(-\delta)(C_{t+1}/C_t)^{\alpha-1}(f(X_{t+1})/f(X_t)), \tag{3.7}$$

where f is some positive function and the Markov state vector X_t is modeled as a stationary process. This representation factors the pricing kernel into the usual power utility part and a multiplicative part $f(X_{t+1})/f(X_t)$ that models the pricing kernel's departure from the permanent consumption growth. This factorization applies to models with habit persistence and to limiting versions of models with recursive utility.[2]

Under this representation, we have

$$\frac{1}{n}E\sum_{j=1}^{n}\log M_{t+j} = -\delta + \frac{(\alpha-1)}{n}E(\log c_{t+n} - \log c_t) + \frac{1}{n}E(\log f(X_{t+n}) - \log f(X_t)). \tag{3.8}$$

[2]See Borovicka, Hansen, Hendricks and Scheinkman (2010).

If we further assume that increments in log consumption growth are i.i.d. and given the stationarity of X_t, we have

$$\frac{1}{n} E \sum_{j=1}^{n} \log M_{t+j} = -\delta + (\alpha - 1) E(\log c_1 - \log c_0). \tag{3.9}$$

Notice that in equation (3.8), the part involving f disappears as stationarity of the state vector implies $\frac{1}{n}E(\log f(X_{t+n}) - \log f(X_t)) = 0$. The formula in equation (3.9) suggests that the expectation of the per period log pricing kernel is a linear (affine, if risk aversion $\alpha > 1$) function of the expected consumption growth rate. This is exactly the same as in the power utility case. Hence, the second component in the basic entropy definition cannot reveal information about the transitory component f. As a result, from the perspective of model comparison, the basic entropy only focuses on fitting the time series of bond yields with different maturities. This absence of information about how the transient component f affects alternative assets (i.e., assets other than riskless bonds) is the primary disadvantage in using the basic entropy to disgnose asset pricing models.

This disappearance of temporary variations in entropy is unique to BCZ's definition. Under generalized entropy, temporary variations show up in the second part of the entropy definition. To see this, we make a few assumptions. First, we assume that log consumption growth and the stationary process $f(X_t)$ are independent and normally distributed. Second, we assume that the increments in log consumption growth are i.i.d. Neither of these assumptions is necessary for the illustration of the usefulness of the generalized entropy. We use them to ease exposition. Under these assumptions, the second part in $GEF(M; s)$ scaled by horizon n is

$$\begin{aligned} \frac{1}{ns} \log E(M_{t+1} \dots M_{t+n})^s &= -\delta + (\alpha - 1) E(\log c_1 - \log c_0) + \frac{1}{2} s (\alpha - 1)^2 \cdot \\ & Var(\log c_1 - \log c_0) + \frac{1}{2} s \frac{Var(\log f(X_{t+n}) - \log f(X_t))}{n}. \end{aligned}$$

For BCZ's entropy, the power s is set at zero so neither the variance of the consumption growth nor the variance of the transient component enters the entropy definition. For alternative values of s, both variances show up. Since $f(X_t)$ is stationary, the scaled variance of the transient shock $Var(\log f(X_{t+n}) - \log f(X_t))/n$ goes to zero as horizon n increases. For a finite horizon n, the second part of GEF weights the unconditional variance of the permanent shock and the temporary shock. In essence, generalized entropy encodes high-order moment[3] information of the pricing kernel that is absent in the basic entropy.

The generalized entropy is crucial for the diagnosis of asset pricing models that feature external habits. Popular habit models such as the Campbell-Cochrane model (Campbell and Cochrane, 1999) and the Santos-Veronesi model (Santos and Veronesi, 2001) have similar but different specifications on transitory shocks through the function f. However, these similar specifications of external habits compound dramatically over time for the pricing of growth rate risk and become drastically different over long horizons.[4] To quantify these differences among models, we first need to have a measure in which the specification of f does matter. Clearly, the basic entropy fails to meet this criteria while the generalized entropy is able to.

Stochastic Volatility

Recent asset pricing models feature stochastic volatility, that is, time-variation in the conditional variance of the pricing kernel. Prominent examples include the long-run risk model (Bansal and Yaron, 2004) and its variations.[5] The defining feature of such a model is the fluctuating quantity of risks. Locally, normality is assumed to obtain

[3] This includes the second moment as shown by the example.

[4] See Borovicka, Hansen, Hendricks and Scheinkman (2010) for an illustration of the nonlinear compounding in the Campbell-Cochrane habit model.

[5] See Ai (2010), Drechsler and Yaron (2011) and Zhou and Zhu (2013).

simple characterization.[6] However, if we integrate over time, such models become a more complicated "mixture of normals". In particular, a low probability event under the steady state volatility can happen with a high probability under a high volatility state. Unconditionally, the thin tails of the normal density are enlarged by this mixture structure. It is thus important to see if the entropy or the generalized entropy can capture this mixture distribution.

To see how the generalized entropy captures stochastic volatility, I use a toy consumption-based model that features stochastic volatility. In particular, suppose the log pricing kernel m_{t+1} has the following form:

$$m_{t+1} = C_0 + C_\sigma \sigma_t^2 - \gamma \sigma_t \eta_{t+1} + A_w w_{t+1}, \tag{3.10}$$

where σ_t is the volatility state variable, η_{t+1} is the permanent shock for consumption growth and w_{t+1} is the shock for the volatility process. The model can be thought of as a restricted long-run risks model in which the long-run growth channel is shut off.[7] In this model, only the permanent shock for consumption growth enters the discount factor with stochastic volatility.

To calculate the second component in the generalized entropy, I first condition on current information and then take unconditional expectation.

$$\begin{aligned} \frac{1}{s} \log E[\exp(s \cdot m_{t+1})] &= \frac{1}{s} \log E[\exp(s \cdot m_{t+1}) | \mathcal{F}_t], & (3.11) \\ &= \text{Const.} + \frac{1}{s} \log E[\exp(s \cdot \underbrace{C_\sigma \sigma_t^2}_{\text{Stationary}} + s^2 \cdot \underbrace{[\frac{1}{2} \gamma^2 \sigma_t^2]}_{\text{Permanent}})] & (3.12) \end{aligned}$$

Based on properties of moment generating functions for normally distributed random variables, the conditional moment generating function for the pricing kernel $E(s \cdot$

[6]Discrete time models that assume normally distributed innovations for the volatility process sometimes imply negative values for volatility and are thus inappropriate. Although alternative distributions (e.g., Gamma distribution) can be applied to obtain strict positivity, I use models with normal innovations to ease the exposition.

[7]See Bansal and Yaron (2004) for the relation between the loadings of the pricing kernel and the preference parameters, especially the IES (Intertemporal Elasticity of Substitution).

$m_{t+1}|\mathcal{F}_t)$ includes two parts. One part is the conditional mean that depends on the stationary state variable σ_t^2 and the other part captures the permanent shock η_{t+1} magnified by the stochastic volatility σ_t. The key observation is that these two parts are weighted differently. BCZ's entropy sets s at zero and thus completely misses the permanent shock part.[8] The intuition is that for an agent with log utility, she optimally chooses the growth optimal portfolio. Her maximized utility only reflects the growth rate of the pricing kernel. At the other extreme, when s equals one (risk aversion $\gamma = 1/(1-s) = \infty$), the agent becomes infinitely risk-averse and the conditional entropy reduces to the reciprocal of the risk-free rate. For s values that fall between zero and one, the agent weights the variation from the state variable σ_t^2 and the variation from the permanent shock.

To sum up, generalized entropy is capable of disclosing more information about the pricing kernel than what the basic entropy can. For a time-homogeneous discount factor, it captures moments of transitory shocks and should be useful for differentiating models that have similar specifications on the nonlinear functional that governs the movement of the state vector (e.g., different versions of the habit model). For a discount factor that features stochastic volatility, it weights moments of the permanent and stationary shocks. It remains interesting to see whether asset market data can provide useful information to bear on these generalized entropies.

3.1.4 Horizon Dependence and Conditioning Information

Horizon dependence is not only an empirical fact for financial market data but also a key feature for most asset pricing models. However, BCZ's entropy captures horizon dependence through bond yields only and completely misses other sources of dependency among for time series of discount factors. To see this, notice that the first

[8] As s approaches zero, $\frac{1}{s}\log E[\exp(s \cdot m_{t+1})]$ approaches $E(m_{t+1})$, which misses the information from contemporaneous shocks.

part in BCZ's entropy $EN(M)$ is log of the expectation of the pricing kernel, which is simply minus the continuously compounded return of a riskless bond. The second part, $E(\log M)$, reduces to the sum of the expectation of single period pricing kernels and, by assuming stationarity of the log pricing kernel, equals the expectation of the single period log pricing kernel times the number of periods. On a per period basis, the second part of their definition always equals the expected single period pricing kernel and fails to capture the dependence structure among them.

The story is different for the generalized entropy. I write out the GEF for an n-period pricing kernel as

$$GEF(M; s) = \log E(M_{t+1}M_{t+2} \dots M_{t+n}) - \frac{1}{s} \log E[(M_{t+1}M_{t+2} \dots M_{t+n})^s]. \quad (3.13)$$

When the discount factors are independent, GEF for the n-period pricing kernel becomes n times the GEF of the single period pricing kernel. In other words, GEF scales in the case of independence. In general, a multi-period GEF cannot be simplified and the second part captures the dependence among single period discount factors other than what has already been captured in bond yields.

Another way to study horizon dependence is to directly consider conditioning information. Previous research has used conditioning information to sharpen unconditional mean-variance frontier and asset pricing bounds.[9] I also use conditioning information to augment the asset space[10] and obtain sharper restrictions on the pricing kernel through generalized entropy bounds. However, we face some choices in selecting among the many financial and economic variables that have been documented to predict market returns. As a first step in bringing in predictive variables to diagnose asset pricing models, this paper relies on empirical proxies for the state variables that candidate models assume. This makes sense because they are the

[9]See Bekaert and Liu (2004), Gallant, Hansen and Tauchen (1990) and Ferson and Siegel (2003).

[10]See Brandt and Santa-Clara (2006) for a similar argument on how to use conditioning information to augment the asset space.

primitive economic assumptions for these models to work and bounds implied by them should be the minimal thresholds for any reasonable model to overcome.

3.2 Applications

3.2.1 Data

For the empirical analysis, I use monthly data on the S&P 500 index and the risk-free rate. The riskfree rate is from Kenneth French's online data library. When conditioning information is considered and for consumption, I use the per capita consumption growth data from the Bureau of Economic Analysis (BEA). For survey forecasts on GDP growth and growth dispersion, I use both the Livingston Survey and the Blue Chips Economic Indicators. I combine forecasts from the two surveys to obtain a large cross-section of forecasters.

3.2.2 Candidate Models

We use three representative agent models that have been proposed by the recent asset pricing literature. For models that feature external habits, we use the Campbell and Cochrane (CC, 1999) model and the Chan and Kogan (CK, 2002) model. Both habit models tie the current price of risk to past consumption but rely on different functional form of the habit stock. CK uses ratio habit while CC uses difference habit. The key parameters in both models are the persistence for habits. I use Campbell and Cochrane (1999)'s choice of 0.9885 for CC and use BCZ's choice of 0.9 for CK. Had we adopted a lower persistence for CK as in Chan and Kogan (2002), the rejection of CK will be even stronger, as we will see later.

For models that feature stochastic volatility, I use the long-run risks model as in Bansal and Yaron (LRR, 2004) but choose to use the calibration in Bansal, Kiku and Yaron (BKY, 2007). BKY suggest a higher persistence for stochastic volatility than the original calibration in Bansal and Yaron (2004) to better match consumption dynamics. In fact, our overall assessment of LRR is independent of the choice of the calibrations.

I choose not to consider the third strand of models proposed by the literature, namely, the rare disaster models. For models that feature skewed and heavy-tailed jumps, we need to consider returns that also have these properties. Liu (2012) takes on this project to use index option returns to infer the rare disaster distributions.

3.2.3 Unconditional Bounds

Time-scaled Generalized Entropy

To ease the interpretation of our results, I define the time-scaled generalized entropy $M(t, s)$ for a multi-period pricing kernel $M_1 M_2 \dots M_t$ as:

$$M(t, s) = \frac{12}{t} \cdot [-\frac{1}{s} \log E(M_1 M_2 \dots M_t)^s]. \tag{3.14}$$

$M(t, s)$ is the second component in GEF multiplied by $12/t$. We choose to focus on $M(t, s)$ as it is the key ingredient that makes GEF different from the basic entropy. Additionally, it is properly scaled so that everything is in annual terms when we vary the horizon t. In particular, when the pricing kernels are i.i.d., $M(t, s)$ reduces to $12 \cdot [-E \log M_1]$, which is constant across all horizon t. Given an arbitrary return R, $12 \cdot [-E \log M_1]$ is bounded below by $12 \cdot [E \log R]$, which is the annal continuously compounded return. Therefore, $M(t, s)$ has a unit that is commensurate with annual returns and is a constant when there is no predictability in the pricing kernel.

Table 3.1 shows the values of $M(t, s)$ across various powers (s) and horizons (t) for the three candidate models. To see how $M(t, s)$ reveals information about the pricing kernel, we start with the baseline case at $s = 0$. This corresponds to a risk aversion of one. At this level of risk aversion, $M(t, 0)$ is constant across all horizons. As we discussed previously, the basic entropy — in the absence of the part that relates to the riskfree rate — fails to capture horizon dependence. To interpret $M(t, 0)$, the

Table 3.1: **Model implied time-scaled generalized entropy** $M(t,s)$**.** The expression for $M(t,s)$ is given by equation (14). We simulate a long time series (50,000) for each candidate model to calculate the expectation in $M(t,s)$.

	Panel A: Chan and Kogan (CK, 2002)				
Power	1mth	3mth	1yr	2yr	5yr
$s=-3(\gamma=1/4)$	161.39	173.34	168.42	141.85	96.47
$s=-1(\gamma=1/2)$	155.32	152.12	129.52	103.67	48.63
$s=0(\gamma=1)$	6.17	6.17	6.17	6.17	6.17
$s=1/2(\gamma=2)$	4.05	2.74	1.92	-1.35	-2.10
$s=2/3(\gamma=3)$	1.90	1.57	-1.71	-2.59	-3.65
	Panel B: Campbell and Cochrane (CC, 1999)				
Power	1mth	3mth	1yr	2yr	5yr
$s=-3(\gamma=1/4)$	86.34	89.89	113.10	125.12	114.14
$s=-1(\gamma=1/2)$	42.33	42.36	42.51	42.91	44.52
$s=0(\gamma=1)$	21.36	21.36	21.36	21.36	21.36
$s=1/2(\gamma=2)$	10.98	11.02	11.19	11.38	11.89
$s=2/3(\gamma=3)$	7.52	7.57	7.74	7.94	8.47
	Panel C: Basal, Kiku and Yaron (BKY, 2007)				
Power	1mth	3mth	1yr	2yr	5yr
$s=-3(\gamma=1/4)$	111.39	113.52	127.22	144.05	138.74
$s=-1(\gamma=1/2)$	54.96	55.03	55.51	56.49	60.50
$s=0(\gamma=1)$	27.30	27.30	27.30	27.03	27.30
$s=1/2(\gamma=2)$	13.53	13.56	13.64	13.70	13.92
$s=2/3(\gamma=3)$	8.94	8.98	9.08	9.14	9.37

basic entropy bound implies that the maximal allowable mean equity return is 27.3% per annum for LRR, which is more than enough to cover the observed mean equity return in the magnitude of 8-10% per annum.[11] For CK and CC, $M(t,0)$ drops to 6.71% and 21.36%, respectively. Similar to what BCZ found, CK is unable to meet the basic entropy. We will discuss the statistical significance of this result later.

For other s values, we see a large amount of variation in $M(t,s)$ across horizon t. For instance, at $s=-3$, $M(t,-3)$ at the five-year horizon is almost 25% higher

[11]Note that $M(t,s)$ does not adjust for the riskfree rate, so we have the mean raw equity return instead of the equity risk premium.

than $M(t, -3)$ at the three-month horizon. The variation is even higher for the two habit models. When the pricing kernels are i.i.d., there should be no variation for $M(t, s)$ across horizon t. Therefore, the amount of variation measures the degree of departure from the i.i.d. case and is important to the diagnosis of asset pricing models that feature predictability and/or stochastic volatility. Fixing horizon t and by varying power s, $M(t, s)$ reveals the degree of non-normality in the pricing kernel. To see this, notice that $M(t, s)$ is linear in s when the multi-period discount factor is log-normally distributed, i.e.,

$$M(t, s) \propto (1 - s) \frac{Var(M_1 M_2 \dots M_t)}{t}. \tag{3.15}$$

Hence, the curvature of $M(t, s)$ in s measures the degree of departure from normality. To evaluate the curvature of $M(t, s)$ based on Table 1, we first approximate the slope of $M(t, s)$ at a power s by

$$\text{Slope}(s) = \frac{M(t, s + \delta) - M(t, s)}{\delta}. \tag{3.16}$$

We calculate the slope at $s = -3$ by setting $\delta = 2$ and the slope at $s = 1/2$ by setting $\delta = 1/6$. For LRR and at one-month horizon, the slopes at $s = -3$ and $s = 1/2$ are 28.40 and 27.84, respectively. At the five-year horizon, they are 37.15 and 26.76, respectively. Therefore, the difference in slopes at different powers is much larger at long horizons than at short horizons. This agrees with our previous discussion on the normal-mixture structure of LRR. With stochastic volatility, the pricing kernel is approximately unconditionally log-normal at monthly horizons but is far from being log-normal at five-year horizons.

To sum up, the time-scaled generalized entropy $M(t, s)$ discloses important information about the discount factor. In particular, horizon dependence is revealed when we vary the horizon t and non-normality is revealed when we vary the power

s. Both horizon dependence and non-normality are key to the diagnosis of recent asset pricing models. It remains interesting to see if asset market data can provide effective restrictions on $M(t, s)$ to help distinguish candidate models.

Model Diagnosis under Unconditional Bounds

We bring in the market data to test the candidate models unconditionally. In particular, given a multi-period pricing kernel $\widetilde{M} = M_{t+1}M_{t+2}\dots M_{t+n}$ and a gross return $\widetilde{R} = R_{t+1}R_{t+2}\dots R_{t+n}$, we test if the following bound is satisfied:

$$-\frac{1}{s}\log E(\widetilde{M}^s) \geqslant \frac{s-1}{s}\log E(\widetilde{R}^{1-\frac{1}{1-s}}),\ \ s < 1. \tag{3.17}$$

As discussed previously, the equivalent risk aversion for power s is $\gamma = \frac{1}{1-s}$. We examine several values of s. The market return $\widetilde{R}$ can represent almost any observable gross return, with the minor restriction that it remains positive with probability one. To be consistent with the literature, I focus on the S&P 500 return data.[12]

To provide statistical significance for a violation of bounds in Equation (3.17), we rely on bootstrap to generate the "p-value" for an inequality. In particular, to take return autocorrelations into account, I use block-bootstrap to generate the empirical distribution for the return moment on the right-hand side of Equation (3.17).[13] The p-value then calculates how often the inequality holds among the bootstrapped samples. A small p-value indicates a high likelihood of violating the bound.

Table 3.2 shows the results for testing Equation (3.17). When $s = -1$, we see that the model implied moments are consistently higher than the data implied moments,

[12] Alternative assets can provide more information about certain properties of the pricing kernel. For instance, Liu (2012) uses option return data to infer the higher order moment properties of the pricing kernel.

[13] In particular, for an n-period return, we split the historical return series into M/n non-overlapping blocks, where M is the total length of the return series. We then treat these blocks as re-sampling units to perform bootstrapping. Newey-West adjustment gives similar results.

Table 3.2: **Testing unconditional bounds.** The inequality given by equation (17) is tested. "Data" shows the right-hand side of equation (17), with $\widetilde{R}$ being the S&P 500 return. The p-values (in bracket) are generated using block-bootstrap as described in footnote (13).

		Horizon				
Power		1mth	3mth	1yr	2yr	5yr
s= -1	Data	10.01	10.11	10.04	9.78	8.92
($\gamma = 1/2$)	LRR	54.96	55.03	55.51	56.49	60.50
		[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	CC	42.33	42.36	42.51	42.91	44.52
		[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	CK	155.32	152.12	129.52	103.67	48.63
		[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
s= 0	Data	9.13	9.12	9.04	8.85	8.27
($\gamma = 1$)	LRR	27.30	27.30	27.30	27.30	27.30
		[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	CC	21.36	21.36	21.36	21.36	21.36
		[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	CK	**6.17**	**6.17**	**6.17**	**6.17**	**6.17**
		[0.32]	[0.32]	[0.34]	[0.36]	[0.38]
s= 1/2	Data	7.35	7.10	6.81	6.53	6.73
($\gamma = 2$)	LRR	13.53	13.56	13.64	13.70	13.91
		[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	CC	8.98	9.02	9.19	9.38	9.89
		[0.77]	[0.78]	[0.81]	[0.83]	[0.88]
	CK	**4.05**	**2.74**	**1.92**	**-1.35**	**-2.10**
		[0.12]	**[0.08]**	**[0.01]**	**[0.00]**	**[0.00]**
s= 2/3	Data	5.52	4.97	4.14	3.27	4.86
($\gamma = 3$)	LRR	8.94	8.98	9.08	9.14	9.37
		[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	CC	6.52	6.57	6.74	6.94	7.47
		[0.86]	[0.87]	[0.89]	[0.78]	[0.81]
	CK	**1.90**	**1.57**	**-1.71**	**-2.59**	**-3.65**
		[0.17]	[0.11]	**[0.03]**	**[0.00]**	**[0.00]**

i.e., the left-hand side of Equation (3.17) consistently dominates the right-hand side. This dominance is statistically significant, as the p-values are uniformly one, implying that there is no violation for the bootstrapped return moments at all.

At $s = 0$, which corresponds to the basic entropy bound, CK violates the bound in mean. That is, the model implied entropy of 6.17% is not enough to explain the observed mean equity return of around 9% across various horizons. This agrees with BCZ, who also find that CK is unable to generate enough dispersion to cover the equity risk premium. However, this violation does not seem to be statistically significant, as the p-values are all above 30% across all horizons. Therefore, there is no statistical evidence for the rejection of CK at $s = 0$. Neither LRR nor CC violates the basic entropy bound.

Turning to other values of s, we find statistical evidence for the rejection of CK. In particular, when $s = 1/2$ or $2/3$, CK fails the generalized entropy bound in mean and moreover, bootstrapped p-values are mostly below 10% when the horizon is above three months and almost zero for two-year and five-year horizon. The rejection seems to be stronger at longer horizons. Neither LRR nor CC violates the bounds in mean.

Overall, using unconditional bounds, we find strong statistical evidence for the rejection of CK. The rejection seems to be stronger at longer horizons. Borovicka et al. (2011) discuss the mechanism of nonlinear compounding in habit models, especially at long horizons. I provide a tool to characterize horizon dependence through the generalized entropy and evaluate model implications through generalized entropy bounds. It seems that bounds with fractional powers (i.e., risk aversion $\gamma > 1$) are powerful in diagnosing asset pricing models.

3.2.4 Conditional Bounds

Conditioning Variables

For unconditional bounds, we are essentially assuming that investors passively hold the market portfolio. But what about other assets and strategies? Certainly, the larger the asset space, the sharper bounds become. So exactly what other assets should we use? One obvious caveat is that no model is the true model. This means we should not use, say, exotic options to gauge general equilibrium CCAPM's.

I suggest the use of dynamic strategies that rely on model implied state variables.

This is reasonable because these state variables are the very basic assumptions for these models to work. If a model offers a reasonable description of the world and given that market participants reside in this model economy and observe these state variables, they will actively trade in the market by following these state variables. Their maximization behavior in turn provides information about the discount factor.[14]

Corresponding to the three candidate models that we focus on, we have two sets of state variables. For habit models, the habit ratio, which relates to past consumption growth, is the key ingredient for the models. I use the realized average consumption growth rate for the past five years as the state variable for the two habit models, denoted as $\hat{h}_t$. For LRR, both the expected growth rate and growth uncertainty play important roles in driving asset prices. I use survey based economic forecasts to construct empirical proxies for these two state variables. In particular, at each date t, suppose we observe the cross-section of forecasts for annual consumption growth rate $E_t^i(\Delta c_{t+1})_{i=1}^n$. We use the sample mean and standard deviation of the cross-section of forecasts as the proxies for the expected growth rate and growth uncertainty, denoted as $\hat{x}_t$ and $\hat{\sigma}_t$, respectively, i.e.,

$$\hat{x}_t = \frac{1}{n}\sum_{i=1}^{n} E_t^i(\Delta c_{t+1}), \tag{3.18}$$

$$\hat{\sigma}_t = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(E_t^i(\Delta c_{t+1}) - \hat{x}_t)^2}. \tag{3.19}$$

Ideally, we would like to have a large cross-section to increase the precision of both measurements. I therefore follow Colacito, Ghysels and Meng (2013) to combine the forecasts from Livingston and SPF surveys. This limits the data frequency to be

[14] See the utility based interpretation of generalized entropy bounds at the beginning of Section 2.

semi-annual as Livingston forecasts are made every half a year. [15]

Predictability Regressions

To take a first look at how these model implied state variables drive returns, I run predictability regressions for both the realized return and return variance. In particular, starting from the day on which forecasts are made, I calculate the realized semi-annual market return $Ret_{t,t+6}$ and realized variance $RV_{t,t+6}$, which is calculated by integrating over weekly returns. I then project these realized variables onto the three state variables. Table 3.3 shows the regression results.

Consistent with habit models, the past consumption growth $\hat{h}_t$ seems to be positively related to both future returns and return variance. However, the loadings do not seem to be significant and the regression R-squares are less than 5%, which are not impressive at the semi-annual frequency. For LRR, the expected growth proxy $\hat{x}_t$ is negatively related to future returns, which is contrary to the implications of standard LRR models (e.g., Bansal and Yaron, 2004).[16] On the other hand, the uncertainty proxy $\hat{\sigma}_t$ positively forecasts both future returns and return variance, consistent with LRR. Moreover, it seems to have modest predictability for future returns ($R^2 = 4\%$) and strong predictability for realized variance ($R^2 = 14\%$). This finding is consistent with Colacito, Ghysels and Meng (2013), who show that forecast dispersion is a significant predictor of return, both in level and in variance.

Model Diagnosis under Conditional Bounds

Predictive regressions show the abilities of some state variables in forecasting the movements of market returns. We now incorporate them into the agent's opti-

[15] For details on how to combine forecasts, see Colacito, Ghysels and Meng (2013). For other works that use survey forecasts to measure growth uncertainty, see Bansal and Shaliastovich (2010).

[16] Further extensions of the model, possibly through richer specifications on the correlation between the expected growth shock and dividend growth shock, can reconcile this finding. See Jagannathan and Marakani (2011).

Table 3.3: **Predictive regressions for model implied state variables.** $\hat{x}_t$ and $\hat{\sigma}_t$ and the mean and standard deviation of the cross-section of survey forecasts as given by equation (18) and (19), respectively. $\hat{h}_t$ is the average consumption growth rate for the past five years. We project the realized return $Ret_{t,t+6}$ and realized variance $RV_{t,t+6}$ onto current state variables. R^2 reports the adjusted R-square.

	$\hat{x}_t$	$\hat{\sigma}_t$	$\hat{h}_t$	R^2
$Ret_{t,t+6}$	**-0.171**			0.06
	(0.07)			
		0.093		0.04
		(0.08)		
			0.371	0.02
			(0.52)	
	-0.052	0.031		0.07
	(0.03)	(0.03)		
$RV_{t,t+6}$	-0.143			0.05
	(0.12)			
		0.160		0.14
		(0.08)		
			0.048	0.04
			(0.05)	
	-0.097	**0.149**		0.17
	(0.10)	(0.07)		

mization problem and use the maximally achievable utility to sharpen asset pricing bounds, thanks to the utility-based interpretation of generalized entropy bounds discussed previously. I do this in a simple manner. I assume that an investor allocates her wealth between the market portfolio and the riskfree rate at the beginning of each period, after observing an investment signal s_t. Assuming a function F that describes the weight that she puts on the market portfolio, the end of period return over horizon h is given by:

$$R_{t,t+h} = F(s_t) \cdot R^{mkt}_{t,t+h} + (1 - F(s_t)) \cdot R^{bond}_{t,t+h}. \tag{3.20}$$

Furthermore, I postulate that F follows a logit function.[17] That is, given a univariate investment signal s_t, $F(s_t)$ is given by:

$$F(s_t) = \frac{\exp(\alpha_0 + \alpha_1 s_t)}{\exp(\alpha_0 + \alpha_1 s_t) + 1}, \tag{3.21}$$

where α_0 and α_1 are coefficients that need to be estimated. There are at least two advantages in adopting the response function F. First, although more complicated functional forms can be used, the simple two-parameter logit specification allows a more robust estimate.[18] Second, the logit function guarantees that the weight strictly falls between zero and one, making the composite return always positive. The positivity of returns is a necessary condition for bounds to hold.

An investor with a risk aversion coefficient of γ seeks to maximize her unconditional expected utility $E(R_{t,t+h}^{1-\gamma})/(1-\gamma)$. Given a response function $F(s_t)$, her utility is determined and depends on α_0 and α_1. I search for α_0 and α_1 to maximize the sample counterpart of her utility. Let the estimated coefficients for the response function be $\hat{\alpha}_0$ and $\hat{\alpha}_1$ and the maximized objective function be $U(\hat{\alpha}_0, \hat{\alpha}_1)$. Then the generalized entropy bounds, applying to the dynamic strategy return $R_{t,t+h}$, implies the following inequality:

$$-\frac{1}{s}\log E(M^s) \geqslant \frac{s-1}{s}\log U(\hat{\alpha}_0, \hat{\alpha}_1)(1-\gamma),\ \gamma = \frac{1}{1-s}. \tag{3.22}$$

Similar to what I do under unconditional bounds, I test if the above inequality holds for a given model. To calculate statistical significance, I need to generate the empirical distribution of $U(\hat{\alpha}_0, \hat{\alpha}_1)$. I achieve this through bootstrapping.[19]

[17]For alternative functional forms, see the later discussion on robustness.

[18]For the tradeoff between robustness and flexibility for portfolio choice problems, see Brandt (1999).

[19]I simultaneously bootstrap the historical return series and investment signals. For each bootstrapped sample, I solve the optimization problem and obtain the maximized utility. I do this many times to obtain the empirical distribution for $U(\hat{\alpha}_0, \hat{\alpha}_1)$.

Table 3.4: **Testing conditional bounds using past consumption growth ($\hat{h}_t$).** The inequality given by equation (22) is tested. "Data(Uncond.)" and "Data(Cond.)" show the right-hand side of equation (17) and (22), respectively. The p-values (in bracket) are generated using block-bootstrap as described in footnote (19).

		Horizon			
		3mth	1yr	2yr	5yr
s= 0	Data(Uncond.)	9.12	9.04	8.85	8.27
($\gamma = 1$)	Data(Cond.)	9.58	10.01	9.78	9.91
	LRR	27.30	27.30	27.30	27.30
		[1.00]	[1.00]	[1.00]	[1.00]
	CC	21.36	21.36	21.36	21.36
		[1.00]	[1.00]	[1.00]	[1.00]
s= 1/2	Data(Uncond.)	7.10	6.81	6.53	6.73
($\gamma = 2$)	Data(Cond.)	7.98	7.51	7.16	7.21
	LRR	13.56	13.64	13.70	13.91
		[1.00]	[1.00]	[1.00]	[1.00]
	CC	9.02	9.19	9.38	9.89
		[0.83]	[0.91]	[1.00]	[1.00]
s= 2/3	Data (Uncond.)	4.97	4.14	3.27	4.86
($\gamma = 3$)	Data (Cond.)	5.11	5.02	4.12	4.37
	LRR	8.98	9.08	9.14	9.37
		[1.00]	[1.00]	[1.00]	[1.00]
	CC	6.57	6.74	6.94	7.47
		[0.89]	[0.93]	[1.00]	[1.00]

Table 3.4-3.6 show the results by examining one signal at a time. By conditioning on the volatility signal $\hat{\sigma}_t$, we see the rejection of CC. In particular, when $s = 1/2$ ($\gamma = 2$), CC cannot generate enough entropy at the two-year and five-year horizon. But the results are statistically insignificant. When $s = 2/3$ ($\gamma = 3$), CC is rejected statistically (at 10%) at relatively short horizons. There seems to be no violation for LRR. Therefore, by incorporating conditional information, the data moments (i.e., right-hand side of Equation (17)) are increased and this helps us better screen candidate models. As a result, CC survives the exercise using unconditional bounds but fails when the volatility signal is considered. When either the growth signal $\hat{x}_t$ or the habit signal $\hat{h}_t$ is used, however, no rejection for CC is found. This is due to the limited ability of either signal to generate significant utility gain over the

Table 3.5: **Testing conditional bounds using expected growth ($\hat{x}_t$).** The inequality given by equation (22) is tested. "Data(Uncond.)" and "Data(Cond.)" show the right-hand side of equation (17) and (22), respectively. The p-values (in bracket) are generated using block-bootstrap as described in footnote (19).

		Horizon			
		3mth	1yr	2yr	5yr
s= 0	Data(Uncond.)	9.12	9.04	8.85	8.27
($\gamma = 1$)	Data(Cond.)	12.13	11.89	11.78	11.21
	LRR	27.30	27.30	27.30	27.30
		[1.00]	[1.00]	[1.00]	[1.00]
	CC	21.36	21.36	21.36	21.36
		[1.00]	[1.00]	[1.00]	[1.00]
s= 1/2	Data(Uncond.)	7.10	6.81	6.53	6.73
($\gamma = 2$)	Data(Cond.)	8.98	8.71	8.35	8.11
	LRR	13.56	13.64	13.70	13.91
		[1.00]	[1.00]	[1.00]	[1.00]
	CC	9.02	9.19	9.38	9.89
		[0.56]	[0.76]	[0.75]	[0.88]
s= 2/3	Data (Uncond.)	4.97	4.14	3.27	4.86
($\gamma = 3$)	Data (Cond.)	6.01	5.56	5.32	5.01
	LRR	8.98	9.08	9.14	9.37
		[0.94]	[1.00]	[1.00]	[1.00]
	CC	6.57	6.74	6.94	7.47
		[0.61]	[0.63]	[0.71]	[0.78]

static strategy (i.e., passively holding the market portfolio) and is consistent with the results from predictive regressions.

Overall, we find statistical evidence for the rejection of CC when growth uncertainty is considered. To the contrary, LRR is not rejected. This points to the importance of time-varying economic uncertainty in driving asset market movements. The habit model, while capable of overcoming the static equity risk premium, seems to have a hard time reconciling the utility gain by an active investor that dynamically balances her portfolio based on growth uncertainty.

Robustness

Different assumptions on the response function may affect the results. An alternative functional form that can also restrict $F(s_t)$ to be within zero and one is the probit

Table 3.6: **Testing conditional bounds using growth uncertainty ($\hat{\sigma}_t$).** The inequality given by equation (22) is tested. "Data(Uncond.)" and "Data(Cond.)" show the right-hand side of equation (17) and (22), respectively. The p-values (in bracket) are generated using block-bootstrap as described in footnote (19).

		Horizon			
		3mth	1yr	2yr	5yr
s= 0	Data(Uncond.)	9.12	9.04	8.85	8.27
($\gamma = 1$)	Data(Cond.)	11.39	11.19	11.01	10.37
	LRR	27.30	27.30	27.30	27.30
		[1.00]	[1.00]	[1.00]	[1.00]
	CC	21.36	21.36	21.36	21.36
		[1.00]	[1.00]	[1.00]	[1.00]
s= 1/2	Data(Uncond.)	7.10	6.81	6.53	6.73
($\gamma = 2$)	Data(Cond.)	8.51	8.83	9.40	10.02
	LRR	13.56	13.64	13.70	13.91
		[1.00]	[1.00]	[1.00]	[1.00]
	CC	9.02	9.19	**9.38**	**9.89**
		[0.64]	[0.55]	[0.48]	[0.45]
s= 2/3	Data (Uncond.)	4.97	4.14	3.27	4.86
($\gamma = 3$)	Data (Cond.)	7.23	7.32	7.97	7.81
	LRR	8.98	9.08	9.14	9.37
		[0.87]	[1.00]	[1.00]	[1.00]
	CC	**6.57**	**6.74**	**6.94**	**7.47**
		[0.08]	**[0.09]**	**[0.06]**	[0.34]

function. I redo the above exercise using the probit function and there seems to be no material change in the results.[20]

To isolate the effect of each of the state variables, I use them separately as the conditioning information. However, more significant results (i.e., sharper bounds) can be obtained by combining these signals. In particular, LRR relies on both the expected growth and growth uncertainty and it seems natural to consider them simultaneously as the conditioning information. I combine the two signals in an additive way. More specifically, I assume that the investment signal s_t loads linearly on both state variables and redo the above exercise. As expected, we obtain a sharper

[20]Details on the results are available upon request.

rejection of CC, with a p-value that is around 2% lower across different horizons. The decline in p-value is only marginal, possibly due to the fact that $\hat{x}_t$ and $\hat{\sigma}_t$ are mildly correlated so there is not much independent information provided by each additional signal.[21]

3.3 Conclusion

Leading asset pricing models are successful in explaining several asset market regularities, e.g., first and second moments of the market return and bond yields. Given their equal success along these dimensions, how do we further distinguish them? More importantly, what other return characteristics are also important to bear on these models? I suggest using the generalized entropy to reveal moment information of the pricing kernel and using the generalized entropy bounds as a diagnostic tool to differentiate candidate models.

Under unconditional bound, I find that the optimized utility of an agent with a certain risk aversion provides a high hurdle for models to overcome. In particular, the Chan and Kogan (2002) model fails to meet such a hurdle, not only in mean but also statistically. Adding conditioning information to the analysis, the rejection of the Chan and Kogan (2002) model is even stronger and moreover, we find mild evidence against the Campbell and Cochrane (1999) model. This is because dynamic strategies that depend on growth uncertainty produce utility gains that are too high for the Campbell and Cochrane (1999) model to reconcile.

While model implied state variables are found to be informative in distinguishing candidate models, we can bring in more conditioning variables to further sharpen the restrictions. For instance, variables such as the pd-ratio, earnings-price ratio and past return volatility have been shown to predict the first and/or second moment of the market return. It would be interesting to see if the long-run risks model can meet

[21]The correlation coefficient between $\hat{x}_t$ and $\hat{\sigma}_t$ is 0.47.

the conditional bounds based on these variables. I leave these to future research.

4

. . . and the Cross-Section of Expected Returns

4.1 The Search Process

Our goal is not to catalogue every asset pricing paper ever published. We narrow the focus to papers that propose and test new factors. For example, Sharpe (1964), Lintner (1965) and Mossin (1966) all theoretically proposed (at roughly the same time), a single factor model — the Capital Asset Pricing Model (CAPM). Beginning with Black, Jensen and Scholes (1972), there are hundreds of papers that test the CAPM. We include the theoretical papers as well as the first paper to empirically test the model, in this case, Black, Jensen and Scholes (1972). We do not include the hundreds of papers that test the CAPM in different contexts, e.g., international markets, different time periods. We do, however, include papers, such as Fama and MacBeth (1973) which tests the market factor as well as two additional factors.

Sometimes different papers propose different empirical proxies for the same type of economic risk. Although they may look similar from a theoretical standpoint, we still include them. An example is the empirical proxies for idiosyncratic financial constraints risk. While Lamont, Polk and Saa-Requejo (2001) use the Kaplan and

Zingales (1997) index to proxy for firm-level financial constraint, Whited and Wu (2006) estimate their own constraint index based on the first order conditions of firms' optimization problem. We include both even though they are likely highly correlated.

Since our focus is on factors that can broadly explain asset market return patterns, we omit papers that focus on a small group of stocks or for a short period of time. This will, for example, exclude a substantial amount of empirical corporate finance research that studies event-driven return movements.

Certain theoretical models lack immediate empirical content. Although they could be empirically relevant once suitable proxies are constructed, we choose to exclude them.

With these rules in mind, we narrow our search to generally the top journals in finance, economics and accounting. To include the most recent research, we search for working papers on SSRN. Working papers pose a challenge because there are thousands of them and they are not refereed. We choose a subset of papers that we suspect are in review at top journals or have been presented at top conferences or are due to be presented at top conferences. We end up using 63 working papers. In total, we focus on 312 published works and selected working papers. We catalogue 315 different factors.[1]

Our collection of 315 factors likely under-represents the factor population. First, we generally only consider top journals. Second, we are very selective in choosing only a handful of working papers. Third, and perhaps most importantly, we should be measuring the number of factors tested (which is unobservable) — that is, we do not observe the factors that were tested but failed to pass the usual significance levels and were never published (see Fama (1991)).

[1]As already mentioned, some of these factors are highly correlated. For example, we include two versions of idiosyncratic volatility — where the residual is defined by different time-series regressions.

4.2 Factor Taxonomy

To facilitate our analysis, we group the factors into different categories. We start with two broad categories: "common" and "individual". "Common" means the factor can be viewed as a proxy for a common source of risk. Risk exposure to this factor or its innovations is supposed to help explain cross-sectional return patterns. "Individual" means the factor is specific to the security or portfolio. A good example is Fama and MacBeth (1973). While the beta against the market return is systematic (exposure to a common risk factor), the standard deviation of the market model residual is security specific and hence an idiosyncratic or individual risk. Many of the individual factors we identify are referred to in the literature as "characteristics".

Based on the unique properties of the proposed factors, we further divide the "common" and "individual" groups into finer categories. In particular, we divide "common" into: "financial", "macro", "microstructure", "behavioral", "accounting" and "other". We divide "individual" into the same categories — except we omit the "macro" classification, which is common, by definition. The following table provides further details on the definitions of these sub-categories and gives examples for each.

Table 4.1: **Factor Classification**

Risk type		**Description**	**Examples**
Common (113)	**Financial** (46)	Proxy for aggregate financial market movement, including market portfolio returns, volatility, squared market returns, etc.	Sharpe (1964): market returns; Kraus and Litzenberger (1976): squared market returns
	Macro (40)	Proxy for movement in macroeconomic fundamentals, including consumption, investment, inflation, etc.	Breeden (1979): consumption growth; Cochrane (1991): investment returns
	Microstructure (11)	Proxy for aggregate movements in market microstructure or financial market frictions, including liquidity, transaction costs, etc.	Pastor and Stambaugh (2003): market liquidity; Lo and Wang (2006): market trading volume
	Behavioral (3)	Proxy for aggregate movements in investor behavior, sentiment or behavior-driven systematic mispricing	Baker and Wurgler (2006): investor sentiment; Hirshleifer and Jiang (2010): market mispricing

	Accounting (8)	Proxy for aggregate movement in firm-level accounting variables, including payout yield, cash flow, etc.	Fama and French (1992): size and book-to-market; Da and Warachka (2009): cash flow
	Other (5)	Proxy for aggregate movements that do not fall into the above categories, including momentum, investors' beliefs, etc.	Carhart (1997): return momentum; Ozoguz (2008): investors' beliefs
Individual (202)	**Financial** (61)	Proxy for firm-level idiosyncratic financial risks, including volatility, extreme returns, etc.	Ang, Hodrick, Xing and Zhang (2006): idiosyncratic volatility; Bali, Cakici and Whitelaw (2011): extreme stock returns
	Microstructure (28)	Proxy for firm-level financial market frictions, including short sale restrictions, transaction costs, etc.	Jarrow (1980): short sale restrictions; Mayshar (1981): transaction costs
	Behavioral (3)	Proxy for firm-level behavioral biases, including analyst dispersion, media coverage, etc.	Diether, Malloy and Scherbina (2002): analyst dispersion; Fang and Peress (2009): media coverage
	Accounting (86)	Proxy for firm-level accounting variables, including PE ratio, debt to equity ratio, etc.	Basu (1977): PE ratio; Bhandari (1988): debt to equity ratio
	Other (24)	Proxy for firm-level variables that do not fall into the above categories, including political campaign contributions, ranking-related firm intangibles, etc.	Cooper, Gulen and Ovtchinnikov (2010): political campaign contributions; Edmans (2011): intangibles

Numbers in parentheses represent the number of factors identified. See Table 5 for details.

4.3 Adjusted T-ratios in Multiple Testing

4.3.1 Why Multiple Testing?

Given so many papers have attempted to explain the same cross-section of expected returns,[2] statistical inference should not be based on a "single" test perspective.[3] Our goal is to provide guidance as to the appropriate significance level using a multiple testing framework.

We want to emphasize that there are many forces that make our guidance lenient, that is, a credible case can be made for even lower p-values. We have already

[2]Strictly speaking, different papers study different sample periods and hence focus on "different" cross-sections of expected returns. However, the bulk of the papers we consider have substantial overlapping sample periods. Also, if one believes that cross-sectional return patterns are stationary, then these papers are studying roughly the same cross-section of expected returns.

[3]When just one hypothesis is tested, we use the term "individual test", "single test" and "independent test" interchangeably. The last term should not be confused with any sort of stochastic independence.

mentioned that we only sample a subset of research papers and the "publication bias/hidden tests" issue (i.e. it is difficult to publish a non-result).[4] However, there is another publication bias that is more subtle. In many scientific fields, replication studies routinely appear in top journals. That is, a factor is discovered, and others try to replicate it. In finance and economics, it is very difficult to publish replication studies. Hence, there is a bias towards publishing "new" factors rather than rigorously verifying the existence of discovered factors.

There are two ways to deal with the bias introduced by multiple testing: out-of-sample validation and using a statistical framework that allows for multiple testing.[5] When feasible, out-of-sample testing is the cleanest way to rule out spurious factors. In their study of anomalies, McLean and Pontiff (2013) take the out-of-sample approach. Their results show a degradation of performance of identified anomalies after publication which is consistent with the statistical bias. It is possible that this degradation is larger than they document. In particular, they drop 10 of their 82 anomalies because they could not replicate the in-sample performance of published studies. Given these non-replicable anomalies were not even able to survive routine data revisions, they are likely to be insignificant strategies, either in-sample or out-of-sample. The degradation from the original published "alpha" is 100% for these strategies — which would lead to a higher average rate of degradation for the 82 strategies.

While the out-of-sample approach has many strengths, it has one important drawback: it cannot be used in real time.[6] In contrast to many tests in the physical

[4]See Rosenthal (1979) for one of the earliest and most influential works on publication bias.

[5]Another approach to test factor robustness is to look at multiple asset classes. This approach has been followed in several recent papers, e.g., Frazzini and Pedersen (2012) and Koijen, Moskowitz, Pedersen and Vrugt (2012).

[6]To make real time assessment in the out-of-sample approach, it is common to hold out some data. However, this is not genuine out-of-sample testing as all the data are observable to researchers. A real out-of-sample test needs data in the future.

sciences, we often need years of data to do an out-of-sample test. We pursue the multiple testing framework because it yields immediate guidance on whether a discovered factor is real.

4.3.2 A Multiple Testing Framework

In statistics, multiple testing refers to simultaneous testing more than one hypothesis. The statistics literature was aware of this multiplicity problem at least 50 years ago.[7] Early generations of multiple testing procedures focus on the control of the *family-wise error rate* (see Section 4.3.1). More recently, increasing interest in multiple testing from the medical literature has spurred the development of methods that control the *false-discovery rate* (see Section 4.3.2). Nowadays, multiple testing is an active research area in both the statistics and the medical literature.[8]

Despite the rapid development of multiple testing methods, they have not attracted much attention in the finance literature.[9] Moreover, most of the research that does involve multiple testing focuses on the Bonferroni adjustment, which is known to be too stringent. Our paper aims to fill this gap by systematically introducing the multiple testing framework.

[7]For early research on multiple testing, see Scheffé's method (Scheffé (1959)) for adjusting significance levels in a multiple regression context and Tukey's range test (Tukey (1977)) on pairwise mean differences.

[8]See Shaffer (1995) for a review of multiple testing procedures that control for the *family-wise error rate*. See Farcomeni (2008) for a review that focuses on procedures that control the *false-discovery rate*.

[9]For the literature on multiple testing corrections for data snooping biases, see Sullivan, Timmermann and White (1999, 2001) and White (2000). For research on data snooping and variable selection in predictive regressions, see Foster, Smith and Whaley (1997) and Lynch and Vital-Ahuja (2012). For applications of multiple testing approach in the finance literature, see for example Shanken (1990), Ferson and Harvey (1999), Boudoukh et al. (2007) and Patton and Timmermann (2010). More recently, the *false discovery rate* and its extensions have been used to study technical trading and mutual fund performance, see for example Barras, Scaillet and Wermers (2010), Bajgrowicz and Scaillet (2012) and Kosowski, Timmermann, White and Wermers (2006). Conrad, Cooper and Kaul (2003) point out that data snooping accounts for a large proportion of the return differential between equity portfolios that are sorted by firm characteristics. Bajgrowicz, Scaillet and Treccani (2013) show that multiple testing methods help eliminate a large proportion of spurious jumps detected using conventional test statistics for high-frequency data. Holland, Basu and Sun (2010) emphasize the importance of multiple testing in accounting research.

First, we introduce a hypothetical example to motivate a more general framework. In Table 4.2, we categorize the possible outcomes of a multiple testing exercise. Panel A displays an example of what the literature could have discovered and Panel B notationalizes Panel A to ease our subsequent definition of the general Type I error rate — the chance of making at least one false discovery or the expected fraction of false discoveries.

Table 4.2: **Contingency Table in Testing M Hypotheses.**

Panel A shows a hypothetical example for factor testing. Panel B presents the corresponding notation in a standard multiple testing framework.

Panel A: An Example

	Unpublished	Published	Total
Truly insignificant	500	50	550
Truly significant	100	50	150
Total	600	100	700

Panel B: The Testing Framework

	H_0 not rejected	H_0 rejected	Total
H_0 True	$N_{0\|a}$	$N_{0\|r}$	M_0
H_0 False	$N_{1\|a}$	$N_{1\|r}$	M_1
Total	$M-R$	R	M

Our example in Panel A assumes 100 published factors (denoted as R). Among these factors, suppose 50 are false discoveries and the rest are real ones. In addition, researchers have tried 600 other factors but none of them were found to be significant. Among them, 500 are truly insignificant but the other 100 are true factors. The total number of tests (M) is 700. Two types of mistakes are made in this process: 50 factors are falsely discovered to be true while 100 true factors are buried in unpublished work. Usual statistical control in a multiple testing context aims at reducing "50" or "50/100", the absolute or proportionate occurrence of false discoveries, respectively. Of course, we only observe published factors because factors that are tried and found to be insignificant rarely make it to publication.[10] This poses a challenge since the usual statistical techniques only handle the case where all test results are observable.

Panel B defines the corresponding terms in a formal statistical testing framework. In a factor testing exercise, the typical null hypothesis is that a factor is not significant. Therefore, a factor is insignificant means the null hypothesis is "true". Using "0" ("1") to indicate the null is true (false) and "a" ("r") to indicate acceptance (rejection), we can easily summarize Panel A. For instance, $N_{0|r}$ measures the number of rejections when the null is true (i.e. the number of false discoveries) and $N_{1|a}$ measures the number of acceptances when the null is false (i.e. the number of missed discoveries). To avoid confusion, we try not to use standard statistical language in describing our notation but rather words unique to our factor testing context. The generic notation in Panel B is convenient for us to formally define different types of errors and describe adjustment procedures in subsequent sections.

4.3.3 Type I and Type II Errors

For a single hypothesis test, a value α is used to control Type I error: the probability of finding a factor to be significant when it is not. In a multiple testing framework, restricting each individual test's Type I error rate at α is not enough to control the overall probability of false discoveries. The intuition is that, under the null that all factors are insignificant, it is very likely for an event with α probability to occur when many factors are tested. In multiple hypothesis testing, we need measures of the Type I error that help us simultaneously evaluate the outcomes of many individual tests.

To gain some intuition on plausible measures of Type I error, we return to Panel B of Table 4.2. $N_{0|r}$ and $N_{1|a}$ count the total number of the two types of errors:

[10] Examples of publication of unsuccessful factors include Fama and MacBeth (1973) and Ferson and Harvey (1993). Fama and MacBeth (1973) show that squared beta and standard deviation of the market model residual have an insignificant role in explaining the cross-section of expected returns. Overall, it is rare to publish "non-results" and all instances of published non-results are coupled with significant results for other factors.

$N_{0|r}$ counts false discoveries while $N_{1|a}$ counts missed discoveries. As generalized from single hypothesis testing, the Type I error in multiple hypothesis testing is also related to false discoveries — concluding a factor is "significant" when it is not. But, by definition, we must draw several conclusions in multiple hypothesis testing, and there is a possible false discovery for each. Therefore, plausible definitions of the Type I error should take into account the joint occurrence of false discoveries.

The literature has adopted at least two ways of summarizing the "joint occurrence". One approach is to count the total number of occurrences $N_{0|r}$. $N_{0|r}$ greater than zero suggests incorrect statistical inference for the overall multiple testing problem — the occurrence of which we should limit. Therefore, the probability of event $N_{0|r} > 0$ should be a meaningful quantity for us to control. Indeed, this is the intuition behind the *family-wise error rate* introduced later. On the other hand, when the total number of discoveries R is large, one or even a few false discoveries may be tolerable. In this case, $N_{0|r}$ is no longer a suitable measure; a certain *false discovery proportion* may be more desirable. Unsurprisingly, the expected value of $N_{0|r}/R$ is the focus of *false discovery rate*, the second type of control.

The two aforementioned measures are the most widely used in the statistics literature. Moreover, many other techniques can be viewed as extensions of these measures.[11] We now describe each measure in detail.

Family-wise Error Rate

The *family-wise error rate* (FWER) is the probability of at least one Type I error:

$$\text{FWER} = Pr(N_{0|r} \geqslant 1).$$

[11] Holm (1979) is the first to formally define the *family-wise error rate*. Benjamini and Hochberg (1995) define and study the *false discovery rate*. Alternative definitions of error rate include *per comparison error rate* (Saville, 1990), *positive false discovery rate* (Storey, 2003) and *generalized false discovery rate* (Sarkar and Guo, 2009).

FWER measures the probability of even a single false discovery, irrespective of the total number of tests. For instance, researchers might test 100 factors; FWER measures the probability of incorrectly identifying one or more factors to be significant. Given significance or threshold level α, we explore two existing methods (Bonferroni and Holm's adjustment) to ensure FWER does not exceed α. Even as the number of trials increases, FWER still measures the probability of a single false discovery. This absolute control is in contrast to the proportionate control afforded by the *false discovery rate* (FDR), defined below.

False Discovery Rate

The *false discovery proportion* (FDP) is the proportion of Type I errors:

$$\text{FDP} = \begin{cases} \dfrac{N_{0|r}}{R} & \text{if } R > 0, \\ 0 & \text{if } R = 0. \end{cases}$$

The *false discovery rate* (FDR) is defined as:

$$\text{FDR} = E[FDP].$$

FDR measures the expected proportion of false discoveries among all discoveries. It is less stringent than FWER and usually much less so when many tests are performed.[12] Intuitively, this is because FDR allows $N_{0|r}$ to grow in proportion to R whereas FWER measures the probability of making even a single Type I error.

[12] There is a natural ordering between FDR and FWER. Theoretically, FDR is always bounded above by FWER, i.e., $FDR \leqslant FWER$. To see this, by definition,

$$\begin{aligned} \text{FDR} &= E[\frac{N_{0|r}}{R}|R > 0]Pr(R > 0) \\ &\leqslant E[I_{(N_{0|r} \geqslant 1)}|R > 0]Pr(R > 0) \\ &= Pr((N_{0|r} \geqslant 1) \cap (R > 0)) \\ &\leqslant Pr(N_{0|r} \geqslant 1) = \text{FWER}, \end{aligned}$$

where $I_{(N_{0|r} \geqslant 1)}$ is an indicator function of event $N_{0|r} \geqslant 1$. This implies that procedures that control FWER under a certain significance level automatically control FDR under the same significance level. In our context, a factor discovery criterion that controls FWER at α also controls FDR at α.

Returning to Example A, Panel A shows that a false discovery event has occurred under FWER since $N_{0|r} = 50 \geqslant 1$ and the realized *FDP* is high, $50/100 = 50\%$. This suggests that the *probability* of false discoveries (FWER) and the *expected* proportion of false discoveries (FDR) may be high.[13] The remedy, as suggested by many FWER and FDR adjustment procedures, would be to lower p-value thresholds for these hypotheses. In terms of Panel A, this would turn some of the 50 false discoveries insignificant and push them into the "Unpublished" category. Hopefully the 50 true discoveries would survive this p-value "upgrade" and remain significant, which is only possible if their p-values are relatively large.

On the other hand, Type II errors — the mistake of missing true factors — are also important in multiple hypothesis testing. Similar to Type I errors, both the total number of missed discoveries $N_{1|a}$ and the fraction of missed discoveries among all abandoned tests $N_{1|a}/(M - R)$ are frequently used to measure the severity of Type II errors.[14] Ideally, one would like to simultaneously minimize the chance of committing a Type I error and that of committing a Type II error. In our context, we would like to include as few insignificant factors (i.e., as low a Type I error rate) as possible and simultaneously as many significant ones (i.e., as low a Type II error rate) as possible. Unfortunately, this is not feasible: as in single hypothesis testing, a decrease in the Type I error rate often leads to an increase in the Type II error rate and vice versa. We therefore seek a balance between the two types of errors. A standard approach is to specify a significance level α for the Type I error rate and

[13] Panel A only shows one realization of the testing outcome for a certain testing procedure (e.g., independent tests). To evaluate FWER and FDR, both of which are expectations and hence depend on the underlying joint distribution of the testing statistics, we need to know the population of the testing outcomes. To give an example that is compatible with Example A, we assume that the t-statistics for the 700 hypotheses are independent. Moreover, we assume the t-statistic for a true factor follows a normal distribution with mean zero and variance one, i.e., $\mathcal{N}(0, 1)$; for a false factor, we assume its t-statistic follows $\mathcal{N}(2, 1)$. Under these assumptions about the joint distribution of the test statistics, we find via simulations that FWER is 100% and FDR is 26%, both exceeding 5%.

[14] See Simes (1986) for one example of Type II error in simulation studies and Farcomeni (2008) for another example in medical experiments.

derive testing procedures that aim to minimize the Type II error rate, i.e., maximize power, among the class of tests with Type I error rate at most α.

When comparing two testing procedures that can both achieve a significance level α, it seems reasonable to use their Type II error rates. However, the exact Type II error rate typically depends on a set of unknown parameters and is therefore difficult to assess.[15] To overcome this difficulty, researchers frequently use distance of the actual Type I error rate to some pre-specified significance level as the measure for a procedure's efficiency. Intuitively, if a procedure's actual Type I error rate is strictly below α, we can probably push this error rate closer to α by making the testing procedure less stringent, i.e., higher p-value threshold so there will be more discoveries. In doing so, the Type II error rate is presumably lowered given the inverse relation between the two types of error rates. Therefore, once a procedure's actual Type I error rate falls below a pre-specified significance level, we want it to be as close as possible to that significance level in order to achieve the smallest Type II error rate. Ideally, we would like a procedure's actual Type I error rate to be exactly the same as the given significance level.

Both FWER and FDR are important concepts that have wide applications in many scientific fields. However, based on specific applications, one may be preferred over the other. When the number of tests is very large, FWER controlling procedures tend to become very tough and eventually lead to a very limited number of discoveries, if any. Conversely, FWER control is more desirable when the number of tests is relatively small, in which case more discoveries can be achieved and at the same time trusted. In the context of our paper, it is difficult to judge whether the number of tests in the finance literature is large. First, we are unsure of the true

[15] In single hypothesis testing, a typical Type II error rate is a function of the realization of the alternative hypothesis. Since it depends on unknown parameter values in the alternative hypothesis, it is difficult to measure directly. The situation is exacerbated in multiple hypothesis testing because the Type II error rate now depends on a multi-dimensional unknown parameter vector. See Zehetmayer and Posch (2010) for power estimation in large-scale multiple testing problems.

number of factors that have been tried. Although there are around 300 published ones, hundreds or even thousands of factors might have been constructed and tested. Second, 300 may seem a large number to researchers in finance but is very small compared to the number of tests conducted in medical research.[16] Given this difficulty, we do not take a stand on the relative appropriateness of these two measures but instead provide adjusted p-values for both. Researchers can compare their p-values with these benchmarks to see whether FDR or even FWER is satisfied.

Related to the false discovery rate, recent research by Lehmann and Romano (2005) tries to control the probability of the realized FDP exceeding a certain threshold value, i.e., $P(FDP > \gamma) \leqslant \alpha$, where γ is the threshold FDP value and α is the significance level.[17] Instead of the expected FDP (i.e., the FDR), Lehmann and Romano's method allows one to make a statement concerning the realized FDP, which might be more desirable in certain applications. For example, targeting the realized FDP is a loss control method and seems more appropriate for risk management or insurance. For our asset pricing applications, we choose to focus on the FDR. In addition, it is difficult to tell whether controlling the realized FDP at $\gamma = 0.1$ with a significance level of $\alpha = 0.05$ is more stringent than controlling FDP at $\gamma = 0.2$ with a significance level of $\alpha = 0.01$. While we use the FDR in our application, we provide some details on the FDP methods in the Appendix.

4.3.4 P-value Adjustment: Three Approaches

The statistics literature has developed many methods to control both FWER and FDR.[18] We choose to present the three most well-known adjustments: Bonferroni,

[16] For instance, tens of thousands of tests are performed in the analysis of DNA microarrays. See Farcomeni (2008) for more applications of multiple testing in medical research.

[17] Also see Romano and Shaikh (2006) and Romano, Shaikh and Wolf (2008).

[18] Methods that control FWER include Holm (1979), Hochberg (1988) and Hommel (1988). Methods that control FDR include Benjamini and Hochberg(1995), Benjamini and Liu (1999) and Benjamini and Yekutieli (2001).

Holm, and Benjamini, Hochberg and Yekutieli (BHY). Both Bonferroni and Holm control FWER, and BHY controls FDR. Depending on how the adjustment is implemented, they can be categorized into two general types of corrections: a "Single step" correction equally adjusts each p-value and a "sequential" correction is an adaptive procedure that depends on the entire distribution of p-values. Bonferroni is a single-step procedure whereas Holm and BHY are sequential procedures. Table **??** summarizes the two properties of the three methods.

Table 4.3: **A Summary of p-value Adjustments**

Adjustment type	Single/Sequential	Multiple test
Bonferroni	Single	FWER
Holm	Sequential	FWER
Benjamini, Hochberg and Yekutieli (BHY)	Sequential	FDR

In the usual multiple testing framework, we observe the outcomes of all test statistics, those rejected as well as not rejected. In our context, however, successful factors are more likely to be published and their p-values observed. This missing observations problem is the main obstacle in applying existing adjustment procedures. In the appendix, we propose a new general methodology to overcome this problem. For now, we assume that all tests and their associated p-values are observed and detail the steps for the three types of adjustments.

Suppose there are in total M tests and we choose to set FWER at α_w and FDR at α_d. In particular, we consider an example with the total number of tests $M = 10$ to illustrate how different adjustment procedures work. For our main results, we set both α_w and α_d at 5%. Table 4.4, Panel A lists the t-ratios and the corresponding p-values for 10 hypothetical tests. The numbers in the table are broadly consistent with the magnitude of t-ratios that researchers report for factor significance. Note that all 10 factors will be “discovered” if we test one hypothesis at a time. Multiple testing adjustments will usually generate different results.

Table 4.4: **An Example of Multiple Testing**

Panel A displays 10 t-ratios and their associated p-values for a hypothetical example. Panel B and C explain Holm’s and BHY’s adjustment procedure, respectively. Bold numbers in each panel are associated with significant factors under the specific adjustment procedure of that panel. M represents the total number of tests (10) and $c(M) = \sum_{j=1}^{M} 1/j$. k is the order of p-values from lowest to highest. α_w is the significance level for Bonferroni’s and Holm’s procedure and α_d is the significance level for BHY’s procedure. Both numbers are set at 5%. The threshold t-ratio for Bonferroni is 0.05%, for Holm 0.60% and for BHY 0.85%.

Panel A: 10 Hypothetical t-ratios and Bonferroni “significant” factors

k	1	2	3	4	5	6	7	8	9	10	# of discoveries
t-ratio	1.99	2.63	2.21	3.43	2.17	2.64	4.56	5.34	2.75	2.49	3
p-value (%)	4.66	0.85	2.71	**0.05**	3.00	0.84	**0.00**	**0.00**	0.60	1.28	

Panel B: Holm adjusted p-values and “significant” factors

New order (k)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	# of discoveries
Old order k	8	7	4	9	6	2	10	3	5	1	4
p-value (%)	0.00	0.00	0.05	0.60	0.84	0.85	1.28	2.71	3.00	4.66	
$\alpha_w/(M+1-k)$	**0.50**	**0.56**	**0.63**	**0.71**	0.83	1.00	1.25	1.67	2.50	5.00	

Panel C: BHY adjusted p-values and “significant” factors

New order (k)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	# of discoveries
Old order k	8	7	4	9	6	2	10	3	5	1	
p-value (%)	0.00	0.00	0.05	0.60	0.84	0.85	1.28	2.71	3.00	4.66	6
$(k \cdot \alpha_d)/(M \times c(M))$ $\alpha_d = 5\%$	**0.15**	**0.21**	**0.50**	**0.70**	**0.85**	**1.00**	1.20	1.35	1.55	1.70	

Bonferroni’s Adjustment

Bonferroni’s adjustment is as follows:

- Reject any hypothesis with p-value $\leqslant \frac{\alpha_w}{M}$:

$$p_i^{Bonferroni} = \min[Mp_i, 1].$$

Bonferroni applies the same adjustment to each test. It inflates the original p-value by the number of tests M; the adjusted p-value is compared with the threshold value α_w.

Example 4.4.1 To apply Bonferroni's adjustment to the example in Table 4.4, we simply multiply all the p-values by 10 and compare the new p-values with $\alpha_w = 5\%$. Equivalently, we can look at the original p-values and consider the cutoff of $0.5\%(= \alpha_w/10)$. This leaves the t-ratio of tests 4,7 and 8 as significant.

Using the notation in Panel B of Table 4.2 and assuming M_0 of the M null hypotheses are true, Bonferroni operates as a single step procedure that can be shown to restrict FWER at levels less than or equal to $M_0\alpha_w/M$, without any assumption on the dependence structure of the p-values. Since $M_0 \leqslant M$, Bonferroni also controls FWER at level α_w.[19]

Holm's Adjustment

Sequential methods have recently been proposed to adjust p-values in multiple hypothesis testing. They are motivated by a seminal paper by Schweder and Spjotvoll (1982), who suggest a graphical presentation of the multiple testing p-values. In particular, using N_p to denote the number of tests that have a p-value exceeding p, Schweder and Spjotvoll (1982) suggest plotting N_p against $(1-p)$. When p is not very small, it is very likely that the associated test is from the null hypothesis. In this case, the p-value for a null test can be shown to be uniformly distributed between 0 and 1. It then follows that for a large p and under independence among tests, the expected number of tests with a p-value exceeding p equals $T_0(1-p)$, where T_0 is the number of null hypotheses, i.e., $E(N_p) = T_0(1-p)$. By plotting N_p against $(1-p)$, the graph should be approximately linear with slope T_0 for large p-values. Points on the graph that substantially deviate from this linear pattern should correspond to non-null hypotheses, i.e., discoveries. The gist of this argument — large and small

[19] The number of true nulls M_0 is inherently unknown, so we usually cannot make Bonferroni more powerful by increasing α_w to $\hat{\alpha} = M\alpha_w/M_0$ (note that $M_0\hat{\alpha}/M = \alpha_w$). However some papers, including Schweder and Spjotvoll (1982) and Hochberg and Benjamini (1990), try to improve the power of Bonferroni by estimating M_0. We try to achieve the same goal by using either Holm's procedure which also controls FWER or procedures that control FDR, an alternative definition of Type I error rate.

p-values should be treated differently — have been distilled into many variations of sequential adjustment methods, among which we will introduce Holm's method that controls FWER and BHY's method that controls FDR.

Holm's adjustment is as follows:

- Order the original p-values such that $p_{(1)} \leqslant p_{(2)} \leqslant \cdots p_{(k)} \leqslant \cdots \leqslant p_{(M)}$ and let associated null hypotheses be $H_{(1)}, H_{(2)}, \cdots H_{(k)} \cdots, H_{(M)}$.
- Let k be the minimum index such that $p_{(k)} > \frac{\alpha_w}{M+1-k}$.
- Reject null hypotheses $H_{(1)} \cdots H_{(k-1)}$ (i.e., declare these factors significant) but not $H_{(k)} \cdots H_{(M)}$.

The equivalent adjusted p-value is therefore

$$p_{(i)}^{Holm} = \min[\max_{j \leqslant i}\{(M-j+1)p_{(j)}\}, 1].$$

Holm's adjustment is a step-down procedure: [20] for the ordered p-values, we start from the smallest p-value and go down to the largest one. If k is the smallest index that satisfies $p_{(k)} > \frac{\alpha_w}{M+1-k}$, we will reject all tests whose ordered index is below k.

To explore how Holm's adjustment procedure works, suppose k_0 is the smallest index such that $p_{(k)} > \frac{\alpha_w}{M+1-k}$. This means that for $k < k_0$, $p_{(k)} \leqslant \frac{\alpha_w}{M+1-k}$. In particular, for $k = 1$, Bonferroni = Holm, i.e., $\frac{\alpha_w}{M} = \frac{\alpha_w}{M+1-(k=1)}$; for $k = 2$, $\frac{\alpha_w}{M} < \frac{\alpha_w}{M+1-(k=2)}$, so Holm is less stringent than Bonferroni. Since less stringent hurdles are applied to the second to the $(k_0 - 1)$th p-values, more discoveries are generated under Holm's than Bonferroni's adjustment.

Example 4.4.2 To apply Holm's adjustment to the example in Table 4.4, we first order the p-values in ascending order and try to locate the smallest index k

[20] Viewing small p-values as "up" and large p-values as "down", Holm's procedure is a "step-down" procedure in that it goes from small p-values to large ones. This terminology is consistent with the statistics literature. Of course, small p-values are associated with "large" values of the test statistics.

that makes $p_{(k)} > \frac{\alpha_w}{M+1-k}$. Table 4.4, Panel B shows the ordered p-values and the associated $\frac{\alpha_w}{M+1-k}$'s. Starting from the smallest p-value and going up, we see that $p_{(k)}$ is below $\frac{\alpha_w}{M+1-k}$ until $k = 5$, at which $p_{(5)}$ is above $\frac{\alpha_w}{10+1-7}$. Therefore, the smallest k that satisfies $p_{(k)} > \frac{\alpha_w}{M+1-k}$ is 5 and we reject the null hypothesis for the first four ordered tests (we discover four factors) and fail to reject the null for the remaining six tests. The original labels for the rejected tests are in the second row in Panel B. Compared to Bonferroni, one more factor (9) is discovered, that is, four factors rather than three are significant. In general, Holm's approach leads to more discoveries and all discoveries under Bonferroni are also discoveries under Holm's.

Like Bonferroni, Holm also restricts FWER at α_w without any requirement on the dependence structure of p-values. It can also be shown that Holm is uniformly more powerful than Bonferroni in that tests rejected (factors discovered) under Bonferroni are always rejected under Holm but not vice versa. In other words, Holm leads to at least as many discoveries as Bonferroni. Given the dominance of Holm over Bonferroni, one might opt to only use Holm. We include Bonferroni because it is the most widely used adjustment and a simple single-step procedure.

Benjamini, Hochberg and Yekutieli's Adjustment

Benjamini, Hochberg and Yekutieli (BHY)'s adjustment is as follows:

- As with Holm's procedure, order the original p-values such that $p_{(1)} \leqslant p_{(2)} \leqslant \cdots p_{(k)} \leqslant \cdots \leqslant p_{(M)}$ and let associated null hypotheses be $H_{(1)}, H_{(2)}, \cdots H_{(k)} \cdots$, $H_{(M)}$.

- Let k be the maximum index such that $p_{(k)} \leqslant \frac{k}{M \times c(M)} \alpha_d$.

- Reject null hypotheses $H_{(1)} \cdots H_{(k)}$ but not $H_{(k+1)} \cdots H_{(M)}$.

The equivalent adjusted p-value is defined sequentially as:

$$p_{(i)}^{BHY} = \begin{cases} p_{(M)} & \text{if } i = M, \\ \min[p_{(i+1)}^{BHY}, \frac{M \times c(M)}{i} p_{(i)}] & \text{if } i \leqslant M-1. \end{cases}$$

where, $c(M)$ is a function of the total number of tests M and controls for the generality of the test. We adopt the choice in Benjamini and Yekutieli (2001) and set $c(M)$ at

$$c(M) = \sum_{j=1}^{M} \frac{1}{j},$$

a value at which the procedure works under arbitrary dependence structure among the p-values. We discuss alternative specifications of $c(M)$ shortly.

In contrast to Holm's, BHY's method starts with the largest p-value and goes up to the smallest one. If k is the largest index that satisfies $p_{(k)} \leqslant \frac{k}{M \times c(M)} \alpha_d$, we will reject all tests (discover factors) whose ordered index is below or equal to k. Also, note that α_d (significance level for FDR) is chosen to be a smaller number than α_w (significance level for FWER). The reason for such a choice is discussed in Section 4.6.

To explore how BHY works, let k_0 be the largest index such that $p_{(k)} \leqslant \frac{k}{M \times c(M)} \alpha_d$. This means that for $k > k_0$, $p_{(k)} > \frac{k}{M \times c(M)} \alpha_d$. In particular, we have $p_{(k_0+1)} > \frac{(k_0+1)}{M \times c(M)} \alpha_d$, $p_{(k_0+2)} > \frac{(k_0+2)}{M \times c(M)} \alpha_d$, $\ldots$, $p_{(M)} > \frac{M}{M \times c(M)} \alpha_d$. We see that the (k_0+1)th to the last null hypotheses, not rejected, are compared to numbers smaller than α_d, the usual significance level in single hypothesis testing. By being stricter than single hypothesis tests, BHY guarantees that the *false discovery rate* is below the prespecified significance level under arbitrary dependence structure among the p-values. See Benjamini and Yekutieli (2001) for details on the proof.

Example 4.4.3 To apply BHY's adjustment to the example in Table 4.4, we first order the p-values in ascending order and try to locate the largest index k that satisfies $p_{(k)} \leqslant \frac{k}{M\times c(M)}\alpha_d$. Table 4.4, Panel C shows the ordered p-values and the associated $\frac{k}{M\times c(M)}\alpha_d$'s. Starting from the largest p-value and going down, we see that $p_{(k)}$ is above $\frac{k}{M\times c(M)}\alpha_d$ until $k = 6$, at which $p_{(6)}$ is below $\frac{k}{10\times 2.93}\alpha_d$. Therefore, the smallest k that satisfies $p_{(k)} \leqslant \frac{k}{M\times c(M)}\alpha_d$ is 6 and we reject the null hypothesis for the first six ordered tests and fail to reject for the remaining four tests. In the end, BHY leads to six significant factors (8,7,4,9,6 and 2), three more than Bonferroni and two more than Holm.[21]

Under independence among p-values, we can gain insight into the choice of $p_{(i)}^{BHY}$ by interpreting $p_{(i)}^{BHY}$ as the solution to a post-experiment maximization problem.[22] In particular, assume all individual hypotheses are performed and their p-values collected. It can be shown that $p_{(i)}^{BHY}$ is the solution to the following problem:

$$\begin{aligned} &\textit{Objective} : \text{Choose } \hat{p} \text{ that maximizes the number of discoveries } n(\hat{p}),\\ &\textit{Constraint} : \hat{p}M/n(\hat{p}) \leqslant \alpha_d. \end{aligned}$$

We first interpret the constraint. Under independence and when each hypothesis is tested individually at level $\hat{p}$, the expected number of false discoveries satisfies $E(N_{0|r}) \leqslant \hat{p}M$. Hence, after observing the outcome of the experiment and thus conditional on having $n(\hat{p})$ discoveries, the FDR is no greater than $\hat{p}M/n(\hat{p})$. The constraint therefore requires the post-experiment FDR to satisfy the pre-specified significance level. Under this constraint, the objective is to choose $\hat{p}$ to maximize the number of discoveries. Since the constraint is satisfied for each realized p-value sequence of the experiment, it is satisfied in expectation as well. In sum, $p_{(i)}^{BHY}$ is the

[21]For independent tests, 10/10 are discovered. For BHY, the effective cutoff is 0.85%, for Bonferroni 0.50% and for Holm 0.60%. The cutoffs are all far smaller than the usual 5%.

[22]This interpretation is shown in Benjamini and Hochberg (1995). Under independence, $c(M) \equiv 1$ is sufficient for BHY to work. See our subsequent discussions on the choice of $c(M)$.

optimal cutoff p-value (i.e., maximal number of discoveries) that satisfies the FDR constraint for each outcome of the experiment.

The choice of $c(M)$ determines the generality of BHY's procedure. Intuitively, the larger $c(M)$ is, the more difficult it is to satisfy the inequality $p_{(k)} \leqslant \frac{k}{M \times c(M)}\alpha_d$ and hence there will be fewer discoveries. This makes it easier to restrict *the false discovery rate* below a given significance level since fewer discoveries are made. In the original work that develops the concept of *false discovery rate* and related testing procedures, $c(M)$ is set equal to one. It turns out that under this choice, BHY is only valid when the test statistics are independent or positively dependent.[23] With our choice of $c(M)$, BHY is valid under any form of dependence among the p-values.[24] Note with $c(M) > 1$, this reduces the size of $\frac{k}{M \times c(M)}\alpha_d$ and it is tougher to satisfy the inequality $p_{(k)} \leqslant \frac{k}{M \times c(M)}\alpha_d$. That is, there will be fewer factors found to be significant.

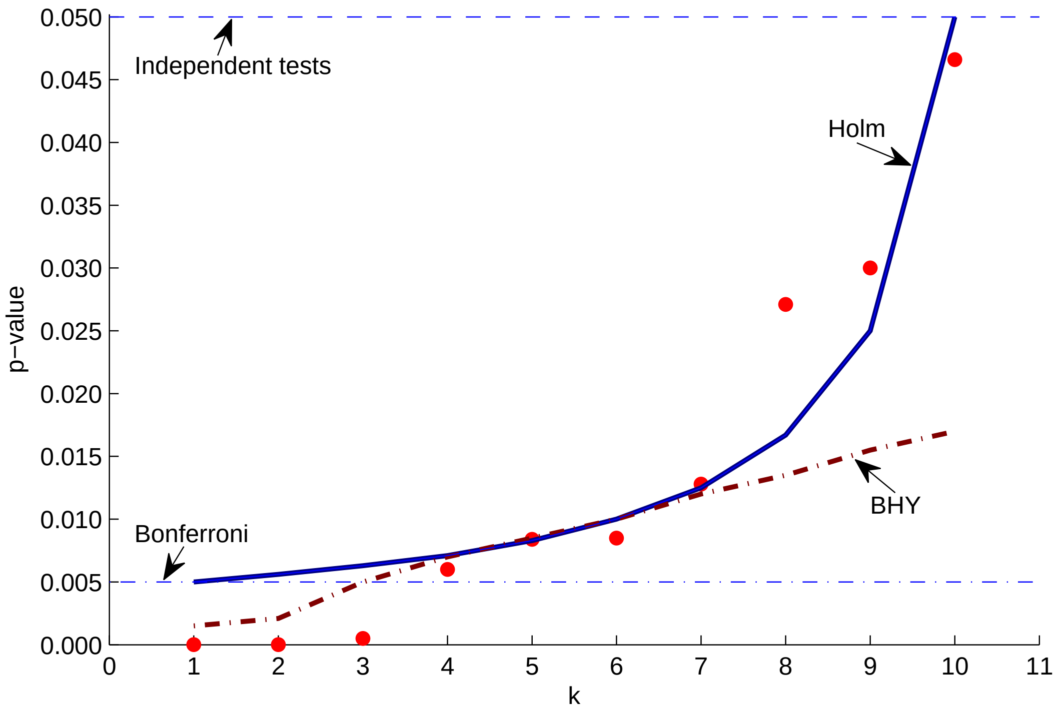
Figure 4.1 summarizes Example A. It plots the original p-value sample as well as threshold p-value lines for various adjustment procedures. We see the stark difference in outcomes between multiple and single hypothesis testing. While all 10 factors would be discovered under single hypothesis testing, only three to six factors would be discovered under a multiple hypothesis test. Although single hypothesis testing guarantees the Type I error of each test meets a given significance level, meeting the more stringent FWER or FDR bound will lead us to discard a number of factors.

To summarize the properties of the three adjustment procedures, Bonferroni's adjustment is the simplest and inflates the original p-value by the total number of

[23]Benjamini and Hochberg (1995) is the original paper that proposes FDR and sets $c(M) \equiv 1$. They show their procedures restricts the FDR below the pre-specified significance level under independence. Benjamini and Yekutieli (2001) and Sarkar (2002) later show that the choice of $c(M) \equiv 1$ also works under positive dependence. For recent studies that assume specific dependence structure to improve on BHY, see Yekutieli and Benjamini (1999), Troendle (2000), Dudoit and Van der Laan (2008) and Romano, Shaikh and Wolf (2008). For a modified Type I error rate definition that is analogous to FDR and its connection to Bayesian hypothesis testing, see Storey (2003).

[24]See Benjamini and Yekutieli (2001) for the proof.

FIGURE 4.1: **Multiple Test Thresholds for Example A**



The 10 p-values for Example A and the threshold p-value lines for various adjustment procedures. All 10 factors are discovered under independent tests, three under Bonferroni, four under Holm and six under BHY. The significance level is set at 5% for each adjustment method.

tests. Holm's adjustment is a refinement of Bonferroni but involves ordering of p-values and thus depends on the entire distribution of p-values. BHY's adjustment, unlike that of Bonferroni or Holm, aims to control the false discovery rate and also depends on the distribution of p-values. Importantly, all three methods allow for general dependence among the test statistics.

4.3.5 Summary Statistics

Figure 4.2 shows the history of discovered factors and publications.[25] We observe a dramatic increase in factor discoveries during the last decade. In the early period

[25]To be specific, we only count those that have t-ratios or equivalent statistics reported. Roughly 20 new factors fail to satisfy this requirement. For details on these, see factors in Table 6 marked with ‡.

from 1980 to 1991, only about one factor is discovered per year. This number has grown to around five in the 1991-2003 period, during which a number of papers, such as Fama and French (1992), Carhart (1997) and Pastor and Stambaugh (2003), spurred interest in studying cross-sectional return patterns. In the last nine years, the annual factor discovery rate has increased sharply to around 18. In total, 162 factors were discovered in the past nine years, roughly doubling the 90 factors discovered in all previous years. We do not include working papers in Figure 4.2. In our sample, there are 63 working papers covering 68 factors.

We obtain t-ratios for each of the 315 factors discovered, including the ones in working papers.[26] The overwhelming majority of t-ratios exceed the 1.96 benchmark for 5% significance.[27] The non-significant ones typically belong to papers that propose a number of factors. These likely represent only a small sub-sample of non-significant t-ratios for all tried factors. Importantly, we take published t-ratios as given. That is, we assume they are econometrically sound with respect to the usual suspects (data errors, coding errors, misalignment, heteroskedasticity, autocorrelation, outliers, etc.).

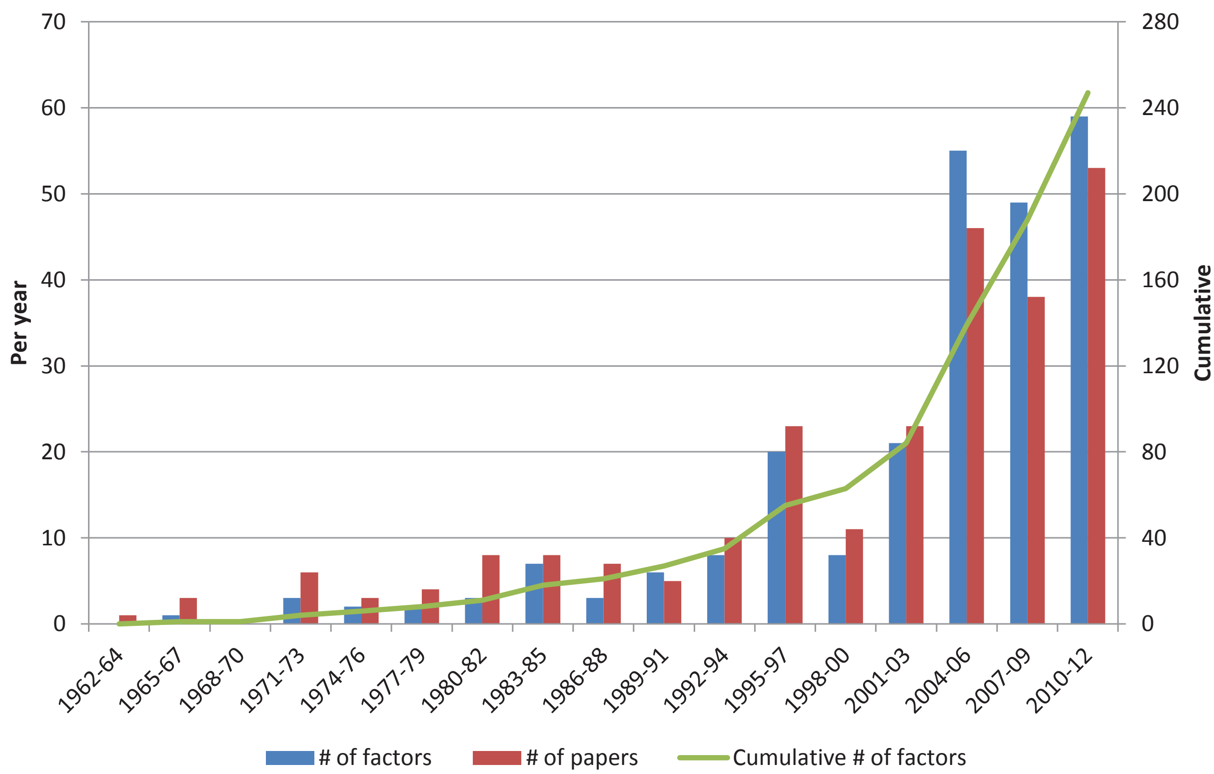
4.3.6 P-value Adjustment When All Tests Are Published ($M = R$)

We now apply the three adjustment methods previously introduced to the observed factor tests, under the assumption that test results of all tried factors are available. We know that this assumption is false since our sample under-represents all

[26]The sign of a t-ratio depends on the source of risk or the direction of the long/short strategy. We usually calculate p-values based on two-sided t-tests, so the sign does not matter. Therefore we use absolute values of these t-ratios.

[27]The multiple testing framework is robust to outliers. The procedures are based on either the total number of tests (Bonferroni) or the order statistics of t-ratios (Holm and BHY). However, outliers may affect our results on the truncated exponential distribution estimation when $M > R$ (see Appendix A). For the results in the paper, we use the full sample. Appendix A shows the implications of trimming the top one percent of t-ratios.

Figure 4.2: **Factors and Publications**



insignificant factors by conventional significance standards: we only observe those insignificant factors that are published alongside significant ones. We design methods to handle this missing data issue later.

Despite some limitations, our results in this section are useful for at least two purposes. First, the benchmark t-ratio based on our incomplete sample provides a lower bound of the true t-ratio benchmark. In other words, if M(total number of tests) $>R$ (total number of discoveries), then we would accept fewer factors than when $M = R$,[28], so future t-ratios need to at least surpass our benchmark to claim

[28]This is always true for Bonferroni's adjustment but not always true for the other two types of adjustments. The Bonferroni adjusted t-ratio is monotonically increasing in the number of trials so the t-ratio benchmark will only rise if there are more factors. Holm and BHY depend on the exact t-ratio distribution so more factors do not necessarily imply a higher t-ratio benchmark.

significance. Second, results in this section can be rationalized within a Bayesian or hierarchical testing framework.[29] Factors in our list constitute an "elite" group: they have survived academia's scrutiny for publication. Placing a high prior on this group in a Bayesian testing framework or viewing this group as a cluster in a hierarchical testing framework, one can interpret results in this section as the first step factor selection within an a priori group.

Based on our sample of observed t-ratios of published factors,[30] we obtain three benchmark t-ratios. In particular, at each point in time, we transform the set of available t-ratios into p-values. We then apply the three adjustment methods to obtain benchmark p-values. Finally, these p-value benchmarks are transformed back into t-ratios, assuming that standard normal distribution well approximates the t-distribution. To guide future research, we extrapolate our benchmark t-ratios 20 years into the future.

We choose to set α_w at 5% (Holm, FWER) and α_d at 1% (BHY, FDR) for our main results. Significance level is subjective, as in individual hypothesis testing where conventional significance levels are usually adopted. Since FWER is a special case of the Type I error in individual testing and 5% seems the default significance level in cross-sectional studies, we set α_w at 5%. On the other hand, FDR is a weaker control relative to FWER; moreover, it has no power in further screening individual

[29]See Wagenmakers and Grünwald (2006) and Storey (2003) on Bayesian interpretations of traditional hypothesis testing. See Meinshausen (2008) for a hierarchical approach on variable selection.

[30]There are at least two ways to generate t-ratios for a risk factor. One way is to show that factor related sorting results in cross-sectional return patterns that are not explained by standard risk factors. The t-ratio for the intercept of the long/short strategy returns regressed on common risk factors is usually reported. The other way is to use factor loadings as explanatory variables and show that they are related to the cross-section of expected returns after controlling for standard risk factors. Individual stocks or stylized portfolios (e.g., Fama-French 25 portfolios) are used as dependent variables. The t-ratio for the factor risk premium is taken as the t-ratio for the factor. In sum, depending on where the new risk factor or factor returns enter the regressions, the first way can be thought of as the left hand side (LHS) approach and the second the right hand side (RHS) approach. For our data collection, we choose to use the RHS t-ratios. When they are not available, we use the LHS t-ratios or simply the t-ratios for the average returns of long/short strategies if the authors do not control for other risk factors.

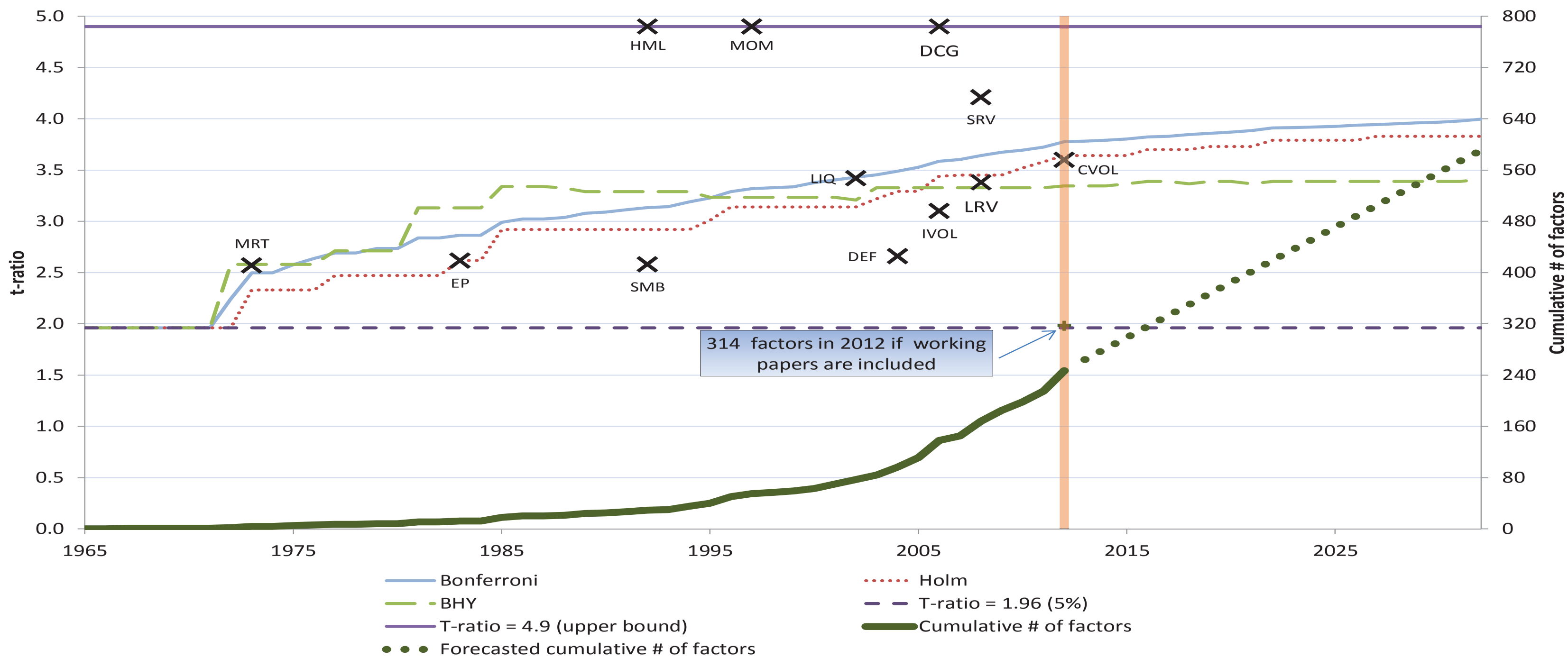
tests if FDR is set greater than or equal to the significance level of individual tests.[31] We therefore set FDR at 1% but will explain what happens when α_d is increased to 5%.

Figure 4.3 presents the three benchmark t-ratios. Both Bonferroni and Holm adjusted benchmark t-ratios are monotonically increasing in the number of discoveries. For Bonferroni, the benchmark t-ratio starts at 1.96 and increases to 3.78 by 2012. It reaches 4.00 in 2032. The corresponding p-values for 3.78 and 4.00 are 0.02% and 0.01% respectively, much lower than the starting level of 5%. Holm implied t-ratios always fall below Bonferroni t-ratios, consistent with the fact that Bonferroni always results in fewer discoveries than Holm. However, Holm tracks Bonferroni closely and their differences are small. BHY implied benchmarks, on the other hand, are not monotonic. They fluctuate before year 2000 and stabilize at 3.39 (p-value = 0.07%) after 2010. This stationarity feature of BHY implied t-ratios, inherent in the definition of FDR, is in contrast to Bonferroni and Holm. Intuitively, at any fixed significance level α, the Law of Large Numbers forces the false discovery rate (FDR) to converge to a constant.[32] If we change α_d to 5%, the corresponding BHY implied benchmark t-ratio is 2.78 (p-value = 0.54%) in 2012 and 2.81 (p-value = 0.50%) in 2032, still much higher than the 1.96 staring value. In sum, taking into account of testing multiplicity, we believe the minimum threshold t-ratio for 5% significance is about 2.8, which corresponds to a p-value of 0.5%.

To see how the new t-ratio benchmarks better differentiate the statistical significance of factors, in Figure 4.3 we mark the t-ratios of a few prominent factors.

[31] When tests are all significant based on single testing and for Benjamini and Hochberg (1995)'s original adjustment algorithm (i.e., $c(M) \equiv 1$), BHY yields the same results as single testing. To see this, notice that the threshold for the largest p-value becomes α_d in BHY's method. As a result, if all tests are individually significant at level α_d, the largest p-value would satisfy $p_{(M)} \leqslant \alpha_d$. Based on BHY's procedure, this means we reject all null hypotheses. In our context, the p-values for published factors are all below 5% due to hidden tests. Therefore, under $c(M) \equiv 1$, if we set α_d equal to 5%, all of these factors will be still be declared as significant under multiple testing.

[32] This intuition is precise for the case when tests are independent. When there is dependence, we need the dependence to be weak to apply the Law of Large Numbers.



The green solid curve shows the historical cumulative number of factors discovered, excluding those from working papers. Forecasts (dotted green line) are based on a linear extrapolation. The dark crosses mark selected factors proposed by the literature. They are MRT (market beta; Fama and MacBeth (1973)), EP (earnings-price ratio; Basu (1983)), SMB and HML (size and book-to-market; Fama and French (1992)), MOM (momentum; Carhart (1997)), LIQ (liquidity; Pastor and Stambaugh (2003)), DEF (default likelihood; Vassalou and Xing (2004)), IVOL (idiosyncratic volatility; Ang, Hodrick, Xing and Zhang (2006)); DCG (durable consumption goods; Yogo (2006)); SRV and LRV (short-run and long-run volatility; Adrian and Rosenberg (2008)) and CVOL (consumption volatility; Boguth and Kuehn (2012)). T-ratios over 4.9 are truncated at 4.9. For detailed descriptions of these factors, see Table 4.6.

FIGURE 4.3: **Adjusted t-ratios, 1965-2032**

Among these factors, HML, MOM, DCG, SRV and MRT are significant across all types of t-ratio adjustments, EP, LIQ and CVOL are sometimes significant and the rest are never significant.

4.3.7 Robustness

Our three adjustment methods are able to control their Type I error rates (FWER for Bonferroni and Holm; FDR for BHY) under arbitrary distributional assumptions about the test statistics. However, if the test statistics are positively correlated, then all three methods might be conservative in that too few factors are discovered. Then again, counteracting this conservatism is our incomplete coverage of tried factors. By adding factors to our current sample, certain adjusted threshold t-ratios (e.g., Bonferroni) will increase, making our current estimates less conservative. We discuss the dependence issue in this section and address the incomplete coverage issue in the Appendix.

Test statistics dependence

In theory, under independence, Bonferroni and Holm approximately achieve the prespecified significance level α when the number of tests is large.[33] On the other hand, both procedures tend to generate fewer discoveries than desired when there is a certain degree of dependence among the tests. Intuitively, in the extreme case where all tests are the same (i.e., correlation = 1.0), we do not need to adjust at all: FWER is the same as the Type I error rate for single tests. Hence, the usual single

[33] To see this for Bonferroni, suppose tests are independent and all null hypotheses are true. We have

$$\begin{aligned} FWER &= Pr(N_{0|r} \geqslant 1) \\ &= 1 - Pr(N_{0|r} = 0) \\ &= 1 - (1 - \alpha/n)^n \\ &\xrightarrow{n \to \infty} 1 - \exp(-\alpha) \approx \alpha \end{aligned}$$

where n denotes the number of tests. The last step approximation is true when α is small.

hypothesis test is sufficient. Under either independence or positive dependence, the actual Type I error rate of BHY is strictly less than the pre-specified significance level, i.e., BHY is too stringent in that too few factors are discovered.[34]

Having discussed assumptions for the testing methods to work efficiently, we now try to think of scenarios that can potentially violate these assumptions. First, factors that proxy for the same type of risk may be dependent. Moreover, returns of long-short portfolios designed to achieve exposure to a particular type of factor may be correlated. For example, hedge portfolios based on dividend yield, earnings yield and book-to-market are correlated. Other examples include risk factors that reflect financial constraints risk, market-wide liquidity and uncertainty risk. If this type of positive dependence exists among test statistics, all three methods would likely to generate fewer significant factors than desired. There is definitely some dependence in our sample. As mentioned previously, there are a number of factors with price in the denominator which are naturally correlated. Another example is that we count four different idiosyncratic volatility factors. On the other hand, most often factors need to "stand their ground" to be publishable. In the end, if you think we are overcounting at 315, consider taking a haircut to 113 factors (the number of "common" factors). Figure 3 shows that our main conclusions do not materially change. For example, the Holm at 113 factors is 3.29 (p-value = 0.10%) while Holm at 315 factors is 3.64 (p-value = 0.03%).

Second, research studying the same factor but based on different samples will generate highly dependent test statistics. Examples include the sequence of papers studying the size effect. We try to minimize this concern by including, with a few exceptions, only the original paper that proposes the factor. To the extent that our list includes few such duplicate factors, our method greatly reduces the dependence that would be introduced by including all papers studying the same factor but for

[34]See footnote 4.3.4 and the references therein.

different sample periods.

Finally, when dependence among test statistics can be captured by Pearson correlations among contemporaneous strategy returns, we present a new model in Section 5 to systematically incorporate the information in test correlations.

The Case When $M > R$

To deal with the hidden tests issue when $M > R$, we propose in Appendix A a simulation framework to estimate benchmark t-ratios. The idea is to first back out the underlying distribution for the t-statistics of all tried factors; then, to generate benchmark t-ratio estimates, apply the three adjustment procedures to simulated t-statistics samples.[35]

Based on our estimates, 70% of all tried factors are missing. The new benchmark t-ratios for Bonferroni and Holm are estimated to be 3.84 and 3.75, respectively; both slightly higher than when $M = R$. This is as expected because more factors are tried under this framework. The BHY implied t-ratio increases from 3.37 to 3.48 at 1% significance and from 2.79 to 2.96 at 5% significance.[36] In sum, across various scenarios, we think the minimum threshold t-ratio is 2.96, corresponding to BHY's adjustment for $M > R$ at 5% significance. Alternative cases all result in even higher benchmark t-ratios. Please refer to Appendix B for the details.

A Bayesian Hypothesis Testing Framework

We can also study multiple hypothesis testing within a Bayesian framework. One major obstacle of applying Bayesian methods in our context is the unobservability of

[35]The underlying assumption for the model in Appendix B is the independence among t-statistics, which may not be plausible given our previous discussions on test dependence. In that case, our structural model proposed in Section 5 provides a more realistic data generating process for the cross-section of test statistics.

[36]Two sets of results are shown in Appendix B: one based on the original sample and the other on the trimmed sample. We use results based on the trimmed sample as they provide the lower bounds on the benchmark t-ratios.

all tried factors. While we propose new frequentist methods to handle this missing data problem, it is not clear how to structure the Bayesian framework in this context. In addition, the high dimensionality of the problem raises concerns on both the accuracy and the computational burden of Bayesian methods.

Nevertheless, ignoring the missing data issue, we outline a standard Bayesian multiple hypothesis testing framework in Appendix B and explain how it relates to our multiple testing framework. We discuss in detail the pros and cons of the Bayesian approach. In contrast to the frequentist approach, which uses generalized Type I error rates to guide multiple testing, the Bayesian approach relies on the posterior likelihood function and thus contains a natural penalty term for multiplicity. However, this simplicity comes at the expense of having a restrictive hierarchical model structure and independence assumptions that may not be realistic for our factor testing problem. Although extensions incorporating certain forms of dependence are possible, it is unclear what precisely we should do for the 315 factors in our list. In addition, even for the Bayesian approach, final reject/accept decision still involves threshold choice. Finally, as the number of tests becomes large, the Bayesian approach gets computationally challenging.[37] Due to these concerns, we choose not to implement the Bayesian approach and instead discuss it briefly. We leave extensions of the basic Bayesian framework that could possibly alleviate the above concerns to future research.

Methods Controlling the FDP

Instead of FDR, recent research by Lehmann and Romano (2005) develops methods to directly control the realized FDP. In particular, they propose a stepdown method to control the probability of FDP exceeding a threshold value. Since their definition

[37]The calculation of the posterior likelihood function involves multiple integrals. As the number of tests becomes large, simulation approaches such as importance sampling may become unstable in calculating these high-dimensional integrals.

of Type I error (i.e., $P(FDP > \gamma)$ where γ is the threshold FDP value) is different from either FWER or FDR, results based on their methods are not comparable to ours. However, the main conclusion is the same. For instance, when $\gamma = 0.10$ and $\alpha = 0.05$, the benchmark t-ratio is 2.70 (p-value = 0.69%), much lower than the conventional cutoff of 1.96. The details are presented in the appendix.

4.4 Correlation Among Test Statistics

Although the BHY method is robust to arbitrary dependence among test statistics, it does not use any information about the dependence structure. Such information, when appropriately incorporated, can be helpful in making the method more accurate (i.e., less conservative). We focus on the type of dependence that can be captured by Pearson correlation. As one way to generate correlation among test statistics, we focus on the correlation among contemporaneous variables (i.e., factor returns) that constitute the test statistics. This is perhaps the most important source of correlation as contemporaneous returns are certainly affected by the movements of the same set of macroeconomic and market variables. Therefore, in our context, the dependence among test statistics is equivalent to the correlation among strategy returns.

Multiple testing corrections in the presence of correlation has only been considered in the recent statistics literature. Existing methods include bootstrap based permutation tests and direct statistical modeling. Permutation tests resample the entire dataset and construct an empirical distribution for the pool of test statistics.[38] Through resampling, the correlation structure in the data is taken into account and no model is needed. In contrast, direct statistical modeling makes additional distri-

[38] Westfall and Young (1993) and Ge et al. (2003) are the early papers that suggest the permutation resampling approach in multiple testing. Later development of the permutation approach tries to reduce computational burden by proposing efficient alternative approaches. Examples include Lin (2005), Conneely and Boehnke (2007) and Han, Kang and Eskin (2009).

butional assumptions on the data generating process. These assumptions are usually case dependent as different kinds of correlations are more plausible under different circumstances.[39]

In addition, recent research in finance explores bootstrap procedures to assess the statistical significance of individual tests.[40] Most of these studies focus on mutual fund evaluation. They bootstrap the time-series of mutual fund returns and obtain an empirical distribution for the t-ratio for each fund. In contrast, our approach focuses on the joint distribution of the t-ratios, as both FWER and FDR depend on the cross-section of t-ratios. As such, we are able to apply a multiple testing framework to the cross-section of factor tests.

Our data pose a challenge to existing methods both in finance and statistics because we do not always observe the time-series of strategy returns (when a t-ratio is based on long-short strategy returns) or the time-series of slopes in cross-sectional regressions (when a t-ratio is based on the slope coefficients in cross-sectional regressions). Often all we have is the single t-statistic that summarizes the significance of a factor. We propose a novel approach to overcome this missing data problem. It is in essence a direct modeling approach but does not require the full information of the return series based on which the t-statistic is constructed. In addition, our approach is flexible enough to incorporate various kinds of distributional assumptions. We expect it to be a valuable addition to the multiple testing literature, especially when only test statistics are observable.

Our method first proposes a structural model to describe the data generating process for the cross-section of returns. It highlights the key statistical properties for returns in our context and is flexible enough to incorporate various kinds of

[39] See Sun and Cai (2008) and Wei et al. (2009).

[40] See Efron (1979) for the original work in the statistics literature. For recent finance applications, see Kosowski, Timmermann, White, and Wermers (2006), Kosowski, Naik and Teo (2007), Fama and French (2010) and Cao, Chen, Liang and Lo (2013).

dependence. Through the structural model, we link Type I error rates in multiple testing to the few structural parameters in the model. Finally, we estimate the model using the t-statistics for published factors and provide multiple testing adjusted t-ratios based on the estimated structural model.[41]

4.4.1 A Model with Correlations

For each factor, suppose researchers construct a corresponding long-short trading strategy and normalize the return standard deviation to be $\sigma = 15\%$ per year, which is close to the annual volatility of the market index.[42] In particular, let the normalized strategy return in period t for the i-th discovered strategy be $X_{i,t}$. Then the t-stat for testing the significance of this strategy is:

$$T_i = (\sum_{t=1}^{N} X_{i,t}/N)/(\sigma/\sqrt{N}).$$

Assuming joint normality and zero serial correlation for strategy returns, this t-stat has a normal distribution

$$T_i \sim N(\mu_i/(\sigma/\sqrt{N}), 1),$$

where μ_i denotes the population mean of the strategy. The μ_i's are unobservable and hypothesis testing under this framework amounts to testing $\mu_i > 0$. We assume that each μ_i is an independent draw from the following mixture distribution:

$$\mu_i \sim p_0 I_{\{\mu=0\}} + (1 - p_0)\text{Exp}(\lambda),$$

where $I_{\{\mu=0\}}$ is the distribution that has a point mass at zero, $\text{Exp}(\lambda)$ is the exponential distribution that has a mean parameter λ and p_0 is the probability of

[41] See Harvey and Liu (2014b) for further details of our approach.

[42] Notice that this assumption is not necessary for our approach. Fixing the standard deviations of different strategies eliminates the need to separately model them, which can be done through a joint modeling of the mean and variance of the cross-section of returns. See Harvey and Liu (2014b) for further discussions on this.

drawing from the point mass distribution. This mixture distribution assumption is the core component for Bayesian multiple testing[43] and succinctly captures the idea of hypothesis testing in the traditional frequentist's view: while there are a range of possible values for the means of truly profitable strategies, a proportion of strategies should have a mean that is indistinguishable from zero. The exponential assumption is not essential for our model as more sophisticated distributions (e.g., a Gamma distribution featuring two free parameters) can be used. We use the exponential distribution for its simplicity[44] and perhaps more importantly, for it being consistent with the intuition that more profitable strategies are less likely to exist. An exponential distribution captures this intuition by having a monotonically decreasing probability density function.

Next, we incorporate correlations into the above framework. Among the various sources of correlations, the cross-sectional correlations among contemporaneous returns are the most important for us to take into account. This is because, unlike time-series correlations for individual return series, cross-sectional return correlations are caused by macroeconomic or market movements and can have a significant impact on multiple testing correction. Other kinds of correlations can be easily embedded into our framework as well.[45]

As a starting point, we assume that the contemporaneous correlation between

[43]See Appendix B for a brief discussion on the Bayesian approach for multiple testing.

[44]As shown later, we need to estimate the parameters in the mixture model based on our t-statistics sample. An over-parameterized distribution for the continuous distribution in the mixture model, albeit flexible, may result in imprecise estimates. We therefore use the simple one-parameter exponential distribution family.

[45]To incorporate the serial correlation for individual strategies, we can model them as simple autoregressive processes. To incorporate the spatial structure in the way that factors are discovered (i.e., a group of factors discovered during a certain period can be related to each other due to the increased research intensity on that group for that period), we can impose a Markov structure on the time-series of μ_i's. See Sun and Cai (2008) for an example of spatial dependence for the null hypotheses. Lastly, to accommodate the intuition that factors within a class should be more correlated than factors across classes, we can use a block diagonal structure for the correlation matrix for strategy returns. See Harvey and Liu (2014b) for further discussion of the kinds of correlation structures that our model is able to incorporate.

two strategies' returns is ρ. The non-contemporaneous correlations are assumed to be zero. That is,

$$\begin{aligned} \text{Corr}(X_{i,t}, X_{j,t}) &= \rho, \quad i \neq j, \\ \text{Corr}(X_{i,t}, X_{j,s}) &= 0, \quad t \neq s. \end{aligned}$$

Finally, to incorporate the impact of hidden tests, we assume that M factors are tried but only factors that exceed a certain t-ratio threshold are published. We set the threshold t-statistic at 1.96 and focus on the sub-sample of factors that have a t-statistic larger than 1.96. However, as shown in Appendix B, factors with marginal t-ratios (i.e., t-ratios just above 1.96) are less likely to be published than those with larger t-ratios. Therefore, our sub-sample of published t-ratios only covers a fraction of t-ratios above 1.96 for tried factors. To overcome this missing data problem, we assume that our sample covers a fraction r of t-ratios in between 1.96 and 2.57 and that all t-ratios above 2.57 are covered. We bootstrap from the existing t-ratio sample to construct the full sample. For instance, when $r = 1/2$, we simply duplicate the sample of t-ratios in between 1.96 and 2.57 and maintain the sample of t-ratios above 2.57 to construct the full sample. For the baseline case, we set $r = 1/2$, consistent with the analysis in Appendix B. We try alternative values of r to see how the results change.[46]

Given the correlation structure and the sampling distribution for the means of returns, we can fully characterize the distributional properties of the cross-section of returns. We can also determine the distribution for the cross-section of t-statistics as they are functions of returns. Based our sample of t-statistics for published research,

[46] Our choice of the threshold t-ratio is smaller than the 2.57 threshold in Appendix A. This is for model identification purposes. With a large t-ratio threshold (e.g., t= 2.57), factors that are generated under the null hypothesis (i.e., false discoveries) are observed with a low probability as their t-ratios rarely exceed the threshold. With little presence of these factors in the sample, certain parameters (e.g., p_0) are poorly identified. In short, we cannot estimate the probability of drawing from the null hypothesis accurately if we rarely observe a factor that is generated from the null hypothesis. We therefore lower the threshold to allow a better estimation of the model. For more details on the selection of the threshold t-ratio, see Harvey and Liu (2014b).

we match key sample statistics with their population counterparts in the model.

The sample statistics we choose to match are the quantiles of the sample of t-statistics and the sample size (i.e., the total number of discoveries). Two concerns motivate us to use quantiles. First, sample quantiles are less susceptible to outliers compared to means and other moment-related sample statistics. Our t-ratio sample does have a few very large observations and we expect quantiles to be more useful descriptive statistics than the mean and the standard deviation. Second, simulation studies show that quantiles in our model are more sensitive to changes in parameters than other statistics. To offer a more efficient estimation of the model, we choose to focus on quantiles.

In particular, the quantities we choose to match and their values for the baseline sample (i.e., $r = 1/2$) are given by:

$$\begin{cases} \widehat{T} = \text{Total number of discoveries} = 352, \\ \widehat{Q}_1 = \text{The 20th percentile of the sample of t-statistics} = 2.39, \\ \widehat{Q}_2 = \text{The 50th percentile of the sample of t-statistics} = 3.16, \\ \widehat{Q}_3 = \text{The 90th percentile of the sample of t-statistics} = 6.35. \end{cases}$$

These three quantiles are representative of the spectrum of quantiles and can be shown to be most sensitive to parameter changes in our model. Fixing the model parameters, we can also obtain the model implied sample statistics T, Q_1, Q_2, and Q_3 through simulations.[47] The estimation works by seeking to find the set of parameters that minimizes the following objective function:

$$D(\lambda, p_0, M, \rho) = w_0(T - \widehat{T})^2 + \sum_{i=1}^{3} w_i(Q_i - \widehat{Q}_i)^2$$

where w_0 and $\{w_i\}_{i=1}^3$ are the weights associated with the squared distances. Mo-

[47]Model implied quantiles are difficult (and most likely infeasible) to calculate analytically. We obtain them through simulations. In particular, for a fixed set of parameters, we simulate 5,000 independent samples of t-statistics. For each sample, we calculate the four summary statistics. The median of these summary statistics across the 5,000 simulations are taken as the model implied statistics.

tivated by the optimal weighting for the Generalized Method of Moments (GMM) estimators, we set these weights at $w_0 = 1$ and $w_1 = w_2 = w_3 = 10,000$. They can be shown to have the same magnitude as the inverses of the variances of the corresponding model implied sample statistics across a wide range of parameter values and should help improve estimation efficiency.

We estimate the three parameters (λ, p_0, and M) in the model and choose to calibrate the correlation coefficient ρ. In particular, for a given level of correlation ρ, we numerically search for the model parameters (λ, p_0, M) that minimize the objective function $D(\lambda, p_0, M, \rho)$.

We choose to calibrate the amount of correlation because the correlation coefficient is likely to be weakly identified in this framework. Ideally, to have a better identification of ρ, we would like to have t-statistics that are generated from samples that have varying degrees of overlap.[48] We do not allow this in either our estimation framework (i.e., all t-statistics are generated from samples that cover the same period) or our data (we do not record the specific period for which the t-statistic is generated). As a result, our results are best interpreted as the estimated t-ratio thresholds for a hypothetical level of correlation. Nonetheless, we provide a brief discussion on the plausible levels of correlation in later sections. For additional details about the estimation method and its performance, we refer the readers to Harvey and Liu (2014b).

To investigate how correlation affects multiple testing, we follow an intuitive simulation procedure. In particular, fixing λ, p_0 and M at their estimates, we know the data generating process for the cross-section of returns. Through simulations, we are able to calculate the previously defined Type I error rates (i.e., FWER and

[48]Intuitively, t-statistics that are based on similar sample periods are more correlated than t-statistics that are based on distinct sample periods. Therefore, the degree of overlap in sample period helps identify the correlation coefficient. See Ferson and Chen (2013) for a similar argument on measuring the correlations among fund returns.

FDR) for any given threshold t-ratio. We search for the optimal threshold t-ratio that exactly achieves a pre-specified error rate.

4.4.2 Results

Our estimation framework assumes a balanced panel with M factors and N periods of returns. We need to assign a value to N. Returns for published works usually cover a period ranging from twenty to fifty years. In our framework, the choice of N does not affect the distribution of T_i under the null hypothesis (i.e., $\mu_i = 0$) but will affect T_i under the alternative hypothesis (i.e., $\mu_i > 0$). When μ_i is different from zero, T_i has a mean of $\mu_i/(\sigma/\sqrt{N})$. A larger N reduces the noise in returns and makes it more likely for T_i to be significant. To be conservative (i.e., less likely to generate significant t-ratios under the alternative hypotheses), we set N at 240 (i.e., twenty years). Other specifications of N change the estimate of λ but leave the other parameters almost intact. In particular, the threshold t-ratios are little changed for alternative values of N.

The results are presented in Table 4.5. Across different correlation levels, λ (the mean parameter for the exponential distribution that represents the mean returns for true factors) is consistently estimated at 0.55% per month. This corresponds to an annual factor return of 6.6%. Therefore, we estimate the average mean returns for truly significant factors to be 6.6% per annum. Given that we standardize factor returns by an annual volatility of 15%, the average annual Sharpe ratio for these factors is 0.44 (or monthly Sharpe ratio of 0.13).

For the other parameter estimates, both p_0 and M are increasing in ρ. Focusing on the baseline case in Panel A and at $\rho = 0$, we estimate that researchers have tried $M = 1,295$ factors and 60.3% ($= 1 - 0.397$) are true discoveries. When ρ is increased to 0.60, we estimate that a total of $M = 1,773$ factors have been tried and around 40.1% ($= 1 - 0.599$) are true factors. Notice that we can estimate the average total number of discoveries by $M \times (1 - p_0)$ if we were able to observe which distribution the factor mean is drawn from. This estimate is around 750 when the level of correlation is not too high (i.e., $\rho < 0.8$). Of course, in reality we cannot observe the underlying distribution for the factor mean and have to rely on the t-statistics. As a result, a significant fraction of these 750 factors are discarded because their associated t-statistics cannot overcome the threshold t-ratio.

Turning to the estimates of threshold t-ratios and focusing on FWER, we see that they are not monotonic in the level of correlation. Intuitively, two forces are at work in driving these threshold t-ratios. On the one hand, both p_0 and M are increasing in the level of correlation. Therefore, more factors — both in absolute value and in proportion — are drawn from the null hypothesis. To control the occurrences of false discoveries based on these factors, we need a higher threshold t-ratio. On the other hand, a higher correlation among test-statistics reduces the required threshold t-ratio. In the extreme case when all test statistics are perfectly correlated, we do not need multiple testing adjustment at all. These two forces work against each other and result in the non-monotonic pattern for the threshold t-ratios under FWER. For FDR, it appears that the impact of larger p_0 and M dominates so that the threshold t-ratios are increasing in the level of correlation.

Across various correlation specifications, our estimates show that in general a t-ratio of 3.9 and 3.0 is needed to control FWER at 5% and FDR at 1%, respectively.[49] Notice that these numbers are not far away from our previous estimates of 3.78 (Holm adjustment that controls FWER at 5%) and 3.38 (BHY adjustment that controls FDR at 1%). However, these seemingly similar numbers are generated through different mechanisms. Our current estimate assumes a certain level of correlation among returns and relies on an estimate of more than 1,300 for the total number of

[49]To save space, we choose not to discuss the performance of our estimation method. Harvey and Liu (2014b) provide a detailed simulation study of our model.

trials. On the other hand, our previous calculation assumes that the 315 published factors are all the factors that have been tried but does not specify a correlation structure.

Table 4.5: **Estimation Results: A Model with Correlations**

We estimate the model with correlations. r is the assumed proportion of missing observations for factors with a t-ratio in between 1.96 and 2.57. Panel A shows the results for the baseline case when $r = 1/2$ and Panel B shows the results for the case when $r = 2/3$. ρ is the correlation coefficient between two strategy returns in the same period. p_0 is the probability of having a strategy that has a mean of zero. λ is the mean parameter of the exponential distribution for the means of the true factors. M is the total number of trials.

				t-ratio			
ρ	p_0	λ(%) (monthly)	M	FWER(5%)	FWER(1%)	FDR(5%)	FDR(1%)
Panel A: $r = 1/2$ (Baseline)							
0	0.397	0.550	1,295	3.89	4.28	2.16	2.88
0.2	0.446	0.555	1,377	3.91	4.30	2.27	2.95
0.4	0.486	0.554	1,476	3.81	4.23	2.34	3.05
0.6	0.599	0.555	1,773	3.67	4.15	2.43	3.09
0.8	0.839	0.560	3,109	3.35	3.89	2.59	3.25
Panel B: $r = 2/3$							
0	0.684	0.550	2,455	4.17	4.55	2.69	3.30
0.2	0.724	0.551	2,696	4.15	4.54	2.76	3.38
0.4	0.774	0.552	3,032	4.06	4.45	2.80	3.40
0.6	0.885	0.562	4,338	3.86	4.29	2.91	3.55
0.8	0.924	0.532	5,391	3.44	4.00	2.75	3.39

4.4.3 How Large Is ρ?

Our sample has limitations in making a direct inference on the level of correlation. To give some guidance, we provide indirect evidence on the plausible levels of ρ.

First, the value of the optimized objective function sheds light on the level of

ρ. Intuitively, a value of ρ that is more consistent with the data generating process should result in a lower optimized objective function. Across the various specifications of ρ in Table 4.5, we find that the optimized objective function reaches its lowest point when $\rho = 0.2$. Therefore, our t-ratio sample suggests a low level of correlation. However, this evidence is only suggestive given the weak identification of ρ in our model.

Second, we draw on external data source to provide inference. In particular, we gain access to the S&P CAPITAL IQ database, which includes detailed information on the time-series of returns of over 400 factors for the US equity market. Calculating the average correlation among these equity risk factors for the 1985-2014 period, we estimate ρ to be around 0.22.

Finally, existing studies in the literature provide guidance on the level of correlation. McLean and Pontiff (2013) estimate the correlation among anomaly returns to be around 0.05. Green, Hand and Zhang (2012) focus on accounting-based factors and find the average correlation to be between 0.06 and 0.20. Focusing on mutual fund returns, Barras, Scaillet and Wermers (2010) argue for a correlation of zero among fund returns while Ferson and Chen (2013) calibrate this number to be between 0.04 and 0.09.

Overall, we believe that the average correlation among factor returns should be low, possibly in the neighborhood of 0.20.

4.5 Conclusion

At least 315 factors have been tested to explain the cross-section of expected returns. Most of these factors have been proposed over the last ten years. Indeed, Cochrane (2011) refers to this as "a zoo of new factors". Our paper argues that it is a serious mistake to use the usual statistical significance cutoffs (e.g., a t-ratio exceeding 2.0) in asset pricing tests. Given the plethora of factors and the inevitable data mining, many of the historically discovered factors would be deemed "significant" by chance.

Our paper presents three conventional multiple testing frameworks and proposes a new one that particularly suits research in financial economics. While these frameworks differ in their assumptions, they are consistent in their conclusions. We argue that a newly discovered factor today should have a t-ratio that exceeds 3.0. We provide a time-series of recommended "cutoffs" from the first empirical test in 1967 through to present day. Many published factors fail to exceed our recommended cutoffs.

While a ratio of 3.0 (which corresponds to a p-value of 0.27%) seems like a very high hurdle, we also argue that there are good reasons to expect that 3.0 is too low. First, we only count factors that are published in prominent journals and we sample only a small fraction of the working papers. Second, there are surely many factors that were tried by empiricists, failed, and never made it to publication or even a working paper. Indeed, the culture in financial economics is to focus on the discovery of new factors. In contrast to other fields such as medical science, it is rare to publish replication studies of existing factors. Given that our count of 315 tested factors is surely too low, this means the t-ratio cutoff is likely even higher.[50]

Should a t-ratio of 3.0 be used for every factor proposed in the future? Probably not. A case can be made that a factor developed from first principles should have a lower threshold t-ratio than a factor that is discovered as a purely empirical exercise. Nevertheless, a t-ratio of 2.0 is no longer appropriate — even for factors that are derived from theory.

In medical research, the recognition of the multiple testing problem has led to the disturbing conclusion that "most claimed research findings are false" (Ioannidis (2005)). Our analysis of factor discoveries leads to the same conclusion – many of the

[50] In astronomy and physics, even higher threshold t-ratios are often used to control for testing multiplicity. For instance, the high profile discovery of Higgs Boson has a t-ratio of more than 5 (p-value less than 0.0001%). See ATLAS Collaboration (2012) and CMS Collaboration (2012).

factors discovered in the field of finance are likely false discoveries: of the 295 published significant factors, 158 would be considered false discoveries under Bonferonni, 142 under Holm, 132 under BHY (1%) and 80 under BHY (5%). In addition, the idea that there are so many factors is inconsistent with the principal component analysis, where, perhaps there are five "statistical" common factors driving time-series variation in equity returns (Ahn, Horenstein and Wang (2012)).

The assumption that researchers follow the rules of classical statistics (e.g., randomization, unbiased reporting, etc.) is at odds with the notion of individual incentives which, ironically, is one of the fundamental premises in economics. Importantly, the optimal amount of data mining is not zero since some data mining produces knowledge. The key, as argued by Glaeser (2008), is to design appropriate statistical methods to adjust for biases, not to eliminate research initiatives. The multiple testing framework detailed in our paper is true to this advice.

Our research quantifies the warnings of both Fama (1991) and Schwert (2003). We attempt to navigate the zoo and establish new benchmarks to guide empirical asset pricing tests.

Table 4.6: **Factor List: Factors Sorted by Year**

An augmented version of this table is available for download and resorting. The main table includes full citations as well as hyperlinks to each of the cited articles. See http://faculty.fuqua.duke.edu/~charvey/Factor-List.xlsx.

Year	Common #	Indi. #	Factor	Formation	Type	Journal	Short reference
1964			Market return	THEORY	Common financial	*Journal of Finance*	Sharpe (1964)
1965			Market return	THEORY	Common financial	*Journal of Finance*	Lintner (1965)
1966			Market return	THEORY	Common financial	*Econometrica*	Mossin (1966)
1967		1	Total volatility	Individual stock return volatility	Individual financial	*Yale Economic Essays*	Douglas (1967)
1972			Market return	THEORY	Common financial	*Journal of Finance*	Heckerman (1972)
			Relative prices of consumption goods	THEORY	Common macro		
1972	1		Market return	Equity index return	Common financial	*Studies in the Theory of Capital Markets*	Black, Jensen and Scholes (1972)
1972			Market return	THEORY	Common financial	*Journal of Business*	Black (1972)
1973			State variables representing future investment opportunity	THEORY	Common financial and macro	*Econometrica*	Merton (1973)
1973			Market return†	Equity index return	Common financial	*Journal of Political Economy*	Fama and MacBeth (1973)
	2		Beta squared*	Square of market beta	Common financial		
		2	Idiosyncratic volatility*	Residual stock volatility from CAPM	Individual financial		
1973			High order market return	THEORY	Common financial	*Journal of Financial and Quantitative Analysis*	Rubinstein (1973)
1974			World market return	THEORY	Common financial	*Journal of Economic Theory*	Solnik (1974)
1974			Individual investor resources	THEORY	Common financial	*Journal of Financial Economics*	Rubinstein (1974)

...continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
1975		3	Earnings growth expectations	Projecting firm earnings growth based on market beta, firm size, dividend payout ratio, leverage and earnings variability	Individual accounting	*Journal of Finance*	Ofer (1975)
1976			Market return†	Equity index return	Common financial	*Journal of Finance*	Kraus and Litzenberger (1976)
	3		Squared market return*	Square of equity index return	Common financial		
1977		4	PE ratio	Firm price-to-earnings ratio	Individual accounting	*Journal of Finance*	Basu (1977)
1978			Marginal rate of substitution	THEORY	Common macro	*Econometrica*	Lucas (1978)
1979		5	Dividend yield	Dividend per share divided by share price	Individual accounting	*Journal of Financial Economics*	Litzenberger and Ramaswamy (1979)
			Market return†	Equity index return	Common financial		
1979			Aggregate real consumption growth	THEORY	Common macro	*Journal of Financial Economics*	Breeden (1979)
1980			Short sale restrictions	THEORY	Individual microstructure	*Journal of Finance*	Jarrow (1980)
1981			Market return†‡	Equity index return	Common financial	*Journal of Finance*	Fogler, John and Tipton (1981)[a]
			Treasury bond return‡	3-month US Treasury bill return	Common financial		
			Corporate bond return‡	Index of long-term Aa utility bonds with deferred calls returns	Common financial		
1981	4		Treasury bill return	Principal components extracted from returns of Treasury bills	Common financial	*Journal of Finance*	Oldfield and Rogalski (1981)
1981			World consumption	THEORY	Common macro	*Journal of Financial Economics*	Stulz (1981)
1981			Transaction costs	THEORY	Individual microstructure	*Journal of Finance*	Mayshar (1981)
1981		6	Firm size	Market value of firm stocks	Individual financial	*Journal of Financial Economics*	Banz (1981)

…continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
1981		7	Short interest	Equity short interest	Individual microstructure	*Journal of Financial and Quantitative Analysis*	Figlewski (1981)
1982			Individual consumer's wealth	THEORY	Common financial	*Journal of Business*	Constantinides (1982)
1983		8	EP ratio	Firm earnings-to-price ratio	Individual accounting	*Journal of Financial Economics*	Basu (1983)
1983			Foreign exchange rate change	THEORY	Common financial	*Journal of Finance*	Adler and Dumas (1983)
1983			Institutional holding‡	Institutional concentration rankings from Standard and Poor's	Individual other	*Financial Analyst Journal*	Arbel, Carvell and Strebel (1983)
1984			Earnings expectations‡	Consensus earnings expectations	Individual accounting	*Financial Analyst Journal*	Hawkins, Chamberlin and Daniel (1984)
1984			New listings announcement‡	Announcement that a company has filed a formal application to list on the NYSE	Individual accounting	*Financial Analyst Journal*	McConnell and Sanger (1984)
1985			Market return†	Equity index return	Common financial	*Journal of Financial Economics*	Chan, Chen and Hsieh (1985)
	5		Industrial production growth	Seasonally adjusted monthly growth rate of industrial production	Common macro		
	6		Change in expected inflation*	Change in expected inflation as defined in Fama and Gibbons (1984)	Common macro		
	7		Unanticipated inflation	Realized minus expected inflation	Common macro		
	8		Credit premium	Risk premium measured as difference in return between "under Baa" bond portfolio and long-term government bond portfolio	Common financial		
	9		Term structure*	Yield curve slope measured as difference in return between long-term government bond and 1-month Treasury bill	Common financial		

...continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
1985		9	Long-term return reversal	Long-term past abnormal return	Individual other	*Journal of Finance*	De Bondt and Thaler (1985)
1985			Investment opportunity change	THEORY	Common financial	*Econometrica*	Cox, Ingersoll and Ross (1985)
1986			Transaction costs	THEORY	Common microstructure	*Journal of Financial Economics*	Amihud and Mendelson (1986)
1986			Transaction costs	THEORY	Common microstructure	*Journal of Political Economy*	Constantinides (1986)
1986			Expected inflation	THEORY	Common macro	*Journal of Finance*	Stulz (1986)
1986	10		Long-term interest rate	Change in the yield of long-term government bonds	Common financial	*Journal of Finance*	Sweeney and Warga (1986)
1986			Industrial production growth†	Seasonally adjusted monthly growth rate of industrial production	Common macro	*Journal of Business*	Chen, Roll and Ross (1986)
			Credit premium†	Risk premium measured as difference in return between "under Baa" bond portfolio and long-term government bond portfolio	Common financial		
			Term structure†	Yield curve slope measured as difference in return between long-term government bond and 1-month Treasury bill	Common financial		
			Unanticipated inflation†	Realized minus expected inflation	Common macro		
			Change in expected inflation†	Changes in expected inflation as defined in Fama and Gibbons (1984)	Common macro		
	11		Change in oil prices*	Growth rate in oil prices	Common macro		
1988		10	Debt to equity ratio	Non-common equity liabilities to equity	Individual accounting	*Journal of Finance*	Bhandari (1988)
1988			Long-term growth forecasts‡	Long-term growth forecasts proxied by the five-year earnings per share growth rate forecasts	Individual accounting	*Financial Analyst Journal*	Bauman and Dowen (1988)

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
1989	12		Consumption growth	Per capita real consumption growth	Common macro	*Journal of Finance*	Breeden, Gibbons and Litzenberger (1989)
1989		11	Illiquidity	Illiquidity proxied by bid-ask spread	Individual microstructure	*Journal of Finance*	Amihud and Mendelson (1989)
1989		12	Predicted earnings change	Predicted earnings change in one year based on a financial statement analysis that combines a large set of financial statement items	Individual accounting	*Journal of Accounting & Economics*	Ou and Penman (1989)
1990		13	Return predictability	Short-term (one month) and long-term (twelve months) serial correlations in returns	Individual financial	*Journal of Finance*	Jegadeesh (1990)
1991			Market return†	Equity index return	Common financial	*Journal of Political Economy*	Ferson and Harvey (1991)
			Consumption growth†	Real per capita growth of personal consumption expenditures for nondurables & services	Common macro		
			Credit spread†	Baa corporate bond return less monthly long-term government bond return	Common financial		
	13		Change in the slope of the yield curve	Change in the difference between a 10-year Treasury bond yield and a 3-month Treasury bill yield	Common financial		
			Unexpected inflation†	Difference between actual and time-series forecasts of inflation rate	Common macro		
	14		Real short rate	One-month Treasury bill return less inflation rate	Common financial		
1992	15		Size	Return on a zero-investment portfolio long in small stocks and short in large stocks	Common accounting	*Journal of Finance*	Fama and French (1992)^b

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
	16		Value	Return on a zero-investment portfolio long in growth stocks and short in value stocks	Common accounting		
1992			Return momentum‡	Size and beta adjusted mean prior five-year returns	Individual financial	*Journal of Financial Economics*	Chopra, Lakonishok and Ritter (1992)
1992			Predicted return signs‡	Return signs predicted by a logit model using financial ratios	Individual accounting	*Journal of Accounting & Economics*	Holthausen and Larcker (1992)
1993		14	Return momentum	Past stock returns	Individual other	*Journal of Finance*	Jegadeesh and Titman (1993)
1993			Returns on S&P stocks‡	Returns on S&P stocks	Common financial	*Review of Financial Studies*	Elton, Gruber, Das and Hlavka (1993)
			Returns on non-S&P stocks‡	Returns on non-S&P stocks	Common financial		
1993			High order market and bond return‡	High order equity index returns and bond returns	Common financial	*Journal of Finance*	Bansal and Viswanathan (1993)[c]
1993			Market return†	Equity index return	Common financial	*Journal of Financial Economics*	Fama and French (1993)
			Size†	Return on a zero-investment portfolio long in small stocks and short in large stocks	Common accounting		
			Value†	Return on a zero-investment portfolio long in growth stocks and short in value stocks	Common accounting		
			Term structure†	Difference in return between long-term government bond and one-month Treasury bill	Common financial		
			Credit risk†	Difference in return between long-term corporate bond and long-term government bond	Common financial		

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
1993			World equity return‡	US dollar return of the MSCI world equity market in excess of a short-term interest rate	Common financial	*Review of Financial Studies*	Ferson and Harvey (1993)^d
			Change in weighted exchange rate‡	Log first difference of the trade-weighted US dollar price of ten industrialized countries' currencies	Common financial		
			Change in long-term inflationary expectations‡	Change in long-term inflationary expectations	Common macro		
			Weighted real short-term interest rate‡	GDP weighted average of short-term interest rates in G-7 countries	Common financial		
			Change in oil price†‡	Change in the monthly average US dollar price per barrel of crude oil	Common macro		
			Change in the Eurodollar-Treasury yield spread‡	First difference of the spread between the 90-day Eurodollar yield and the 90-day Treasury-bill yield	Common financial		
			Change in G-7 industrial production‡	Change in G-7 industrial production	Common macro		
			Unexpected inflation for the G-7 countries‡	Unexpected inflation based on a time-series model on an aggregate G-7 inflation rate	Common macro		
1994	17		World equity return	US dollar return of the MSCI world equity market in excess of a short-term interest rate	Common financial	*Journal of Banking and Finance*	Ferson and Harvey (1994)
	18		Change in weighted exchange rate*	Log first difference of the trade-weighted US dollar price of ten industrialized countries' currencies	Common financial		
	19		Change in long-term inflationary expectations*	Change in long-term inflationary expectations	Common macro		
			Change in oil price*†	Change in the monthly average US dollar price per barrel of crude oil	Common macro		

...continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
1994	20		Tax rate for capital gains	Short-term capital gains tax rate	Common accounting	*Journal of Finance*	Bossaerts and Dammon (1994)
	21		Tax rate for dividend	Dividend tax rate	Common accounting		
1995	22		Change in expected inflation	Change in expectation from economic surveys	Common macro	*Journal of Finance*	Elton, Gruber and Blake (1995)
	23		Change in expected GNP	Change in expectation from economic surveys	Common macro		
1995		15	New public stock issuance	New public stock issuance	Individual accounting	*Journal of Finance*	Loughran and Ritter (1995)
1995		16	Dividend initiations	Initiations of cash dividend payments	Individual financial	*Journal of Finance*	Michaely, Thaler and Womack (1995)
		17	Dividend omissions	Omissions of cash dividend payments	Individual financial		
1995			Seasoned equity offerings‡	Whether a firm makes seasoned equity offerings	Individual financial	*Journal of Financial Economics*	Spiess and Affleck-Graves (1995)
1996	24		Money growth	M2 or M3 minus currency, divided by total population	Common macro	*Journal of Finance*	Chan, Foresi and Lang (1996)
1996	25		Returns on physical investment	Inferred from investment data via a production function	Common macro	*Journal of Political Economy*	Cochrane (1996)
1996			Market return†	Equity index return	Common financial	*Journal of Political Economy*	Campbell (1996)
	26		Labor income	Real labor income growth rate	Common macro		
			Dividend yield†	Dividend yield on value-weighted index	Common financial		
			Interest rate†	Treasury bill rate less 1-year moving average	Common financial		
			Term structure†	Long-short government bond yield spread	Common financial		
1996			Market return†	Equity index return	Common financial	*Journal of Finance*	Jagannathan and Wang (1996)
			Slope of yield curve†	Long-short government bond yield spread	Common financial		

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
			Labor income†	Real labor income growth rate	Common macro		
1996		18	Earnings forecasts	Errors in analysts' forecasts on earnings growth	Individual accounting	*Journal of Finance*	La Porta (1996)
1996		19	R&D capital	R&D capital over total assets	Individual accounting	*Journal of Accounting & Economics*	Lev and Sougiannis (1996)
1996		20	Accruals	Accruals defined by the change in non-cash current assets, less the change in current liabilities, less depreciation expense, all divided by average total assets	Individual accounting	*Accounting Review*	Sloan (1996)
1996		21	Buy recommendations	Buy recommendations from security analysts	Individual financial	*Journal of Finance*	Womack (1996)
		22	Sell recommendations	Sell recommendations from security analysts	Individual financial		
1996		23	Credit rating	Institutional investor country credit rating from semi-annual survey	Individual other	*Journal of Portfolio Management*	Erb, Harvey and Viskanta (1996)
1996		24	Illiquidity	Derivative transaction price with respect to signed trade size	Individual microstructure	*Journal of Financial Economics*	Brennan and Subrahmanyam (1996)
1997			Nonlinear functions of consumption growth‡	Low order orthonormal polynomials of current and future consumption growth	Common macro	*Journal of Finance*	Chapman (1997)[e]
1997			Opportunistic strategy return‡	Return for hedge funds that follow an opportunistic strategy	Common financial	*Review of Financial Studies*	Fung and Hsieh (1997)[f]
			Global/macro strategy return‡	Return for hedge funds that follow a global/macro strategy	Common financial		
			Value strategy return‡	Return for hedge funds that follow a value strategy	Common financial		
			Trend following strategy return‡	Return for hedge funds that follow a trend following strategy	Common financial		

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
			Distressed investment strategy return‡	Return for hedge funds that follow a distressed investment strategy	Common financial		
1997			Size†	Return on a zero-investment portfolio long in small stocks and short in large stocks	Common accounting	*Journal of Finance*	Carhart (1997)
			Value†	Return on a zero-investment portfolio long in growth stocks and short in value stocks	Common accounting		
			Market return†	Equity index return	Common financial		
	27		Momentum	Return on a zero-investment portfolio long in past winners and short in past losers	Common other		
1997			Size†	Market value of equity	Individual accounting	*Journal of Financial Economics*	Brennan, Chordia and Subrahmanyam (1997)
			Book-to-market ratio†	Book value of equity plus deferred taxes to market value of equity	Individual accounting		
			Momentum†	Past cumulative stock return	Individual financial		
		25	Trading volume	Dollar volume traded per month	Individual microstructure		
1997		26	Disclosure level	Voluntary disclosure level of manufacturing firms' annual reports	Individual accounting	*Accounting Review*	Botosan (1997)
1997		27	Earnings forecast uncertainty	Standard deviation of earnings forecasts	Individual accounting	*Journal of Financial Research*	Ackert and Athanassakos (1997)
1997			Size†	Market value of equity	Individual accounting	*Journal of Finance*	Daniel and Titman (1997)
			Value†	Book value of equity plus deferred taxes to market value of equity	Individual accounting		

...continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
1997			Earnings management likelihood‡	Earnings management likelihood obtained by regressiong realized violators of Generally Accepted Accounting Principles on firm characteristics	Individual accounting	*Journal of Accounting and Public Policy*	Beneish (1997)
1997		28	Corporate acquisitions	Difference between stock mergers and cash tender offers for corporate acquisitions	Individual financial	*Journal of Finance*	Loughran and Vijh (1997)
1998			Fundamental analysis‡	Investment signals constructed using a collection of variables that relate to contemporaneous changes in inventories, accounts receivables, gross margins, selling expenses, capital expenditures, effective tax rates, inventory methods, audit qualifications, and labor force sales productivity	Individual accounting	*Accounting Review*	Abarbanell and Bushee (1998)
1998			Firm fundamental value‡	Firms' fundamental values estimated from I/B/E/S consensus forecasts and a residual income model	Individual accounting	*Journal of Accounting and Economics*	Frankel and Lee (1998)
1998		29	Bankruptcy risk	The probability of bankruptcy from Altman (1968)	Individual financial	*Journal of Finance*	Ilia (1998)
1998		30	Illiquidity	Liquidity proxied by the turnover rate: number of shares traded as a fraction of the number of shares outstanding	Individual microstructure	*Journal of Financial Markets*	Datar, Naik and Radcliffe (1998)
1999	28		Fitted return based on predictive regressions	Expected portfolio return obtained by projecting historical returns on lagged macro instruments, including term spreads, dividend yield, credit spread and short-term Treasury bill	Common financial	*Journal of Finance*	Ferson and Harvey (1999)

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
1999		31	Industry momentum	Industry-wide momentum returns	Individual other	*Journal of Finance*	Moskowitz and Grinblatt (1999)
1999			Debt offerings‡	Whether a firm makes straight and convertible debt offerings	Individual financial	*Journal of Financial Economics*	Spiess and Affleck-Graves (1999)
2000	29		Entrepreneur income	Proprietary income of entrepreneurs	Common financial	*Journal of Finance*	Heaton and Lucas (2000)
2000	30		Coskewness	Excess return on a portfolio which long stocks with low past coskewness	Common financial	*Journal of Finance*	Harvey and Siddique (2000)
2000		32	Trading volume	Past trading volume	Individual microstructure	*Journal of Finance*	Lee and Swaminathan (2000)
2000		33	Within-industry size	Difference between firm size and average firm size within the industry	Individual financial	*Working Paper*	Asness, Porter and Stevens (2000)
		34	Within-industry value	Difference between firm book-to-market ratio and average book-to-market ratio within the industry	Individual accounting		
		35	Within-industry cashflow to price ratio	Difference between firm cashflow to price ratio and average cashflow to price ratio within the industry	Individual accounting		
		36	Within-industry percent change in employees	Difference between firm percent change in employees and average percent change in employees within the industry	Individual accounting		
		37	Within-industry momentum	Difference between firm past stock prices and average past stock prices within the industry	Individual financial		
2000		38	Financial statement information	A composite score based on historical financial statement that separates winners from losers	Individual accounting	*Journal of Accounting Research*	Piotroski (2000)
2001			Consumption growth†	Per capita real consumption growth rate	Common macro	*Journal of Political Economy*	Lettau and Ludvigson (2001)

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
	31		Consumption-wealth ratio	Proxied by a weighted average of human and non-human wealth	Common macro		
2001		39	Level of liquidity	Level of dollar trading volume and share turnover	Individual microstructure	*Journal of Financial Economics*	Chordia, Subrahmanyam and Anshuman (2001)
		40	Variability of liquidity	Volatility of dollar trading volume and share turnover	Individual microstructure		
2001		41	Financial constraints	Measure financial constraints with Kaplan and Zingales (1997) index	Individual financial	*Review of Financial Studies*	Lamont, Polk and Saa-Requejo (2001)
2001			Straddle return‡	Lookback straddles' returns constructed based on option prices	Common financial	*Review of Financial Studies*	Fung and Hsieh (2001)
2001			Consensus recommendations*	Consensus recommendations measured by the average analyst recommendations	Individual accounting	*Journal of Finance*	Barber, Lehavy, McNichols and Trueman (2001)
2001		42	Bond rating changes	Moody's bond ratings changes	Individual financial	*Journal of Finance*	Dichev and Piotroski (2001)
2001		43	Analysts' forecasts	Financial analysts' forecasts of annual earnings	Individual accounting	*Accounting Review*	Pieter, Lo and Pfeiffer (2001)
2001		44	Institutional ownership	Institutional holdings of firm assets	Individual accounting	*Quarterly Journal of Economics*	Gompers and Metrick (2001)
2002			Market return†	Equity index return	Common financial	*Journal of Finance*	Dittmar (2002)
			Squared market return†	Squared equity index return	Common financial		
			Labor income growth†	Smoothed labor income growth rate	Common financial		
	32		Squared labor income growth	Squared smoothed labor income growth rate	Common financial		
2002		45	Distress risk	Distress risk as proxied by Ohlson's O-score	Individual financial	*Journal of Finance*	Griffin and Lemmon (2002)
2002		46	Analyst dispersion	Dispersion in analysts' earnings forecasts	Individual behavioral	*Journal of Finance*	Diether, Malloy and Scherbina (2002)

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2002		47	Breadth of ownership	Ratio of the number of mutual funds holding long positions in the stock to total number of mutual funds	Individual microstructure	*Journal of Financial Economics*	Chen, Hong and Stein (2002)
2002		48	Information risk	Probability of information-based trading for individual stock	Individual microstructure	*Journal of Finance*	Easley, Hvidkjaer and O'Hara (2002)
2002		49	Short-sale constraints	Shorting costs for NYSE stocks	Individual microstructure	*Journal of Financial Economics*	Jones and Lamont (2002)
2002		50	Earnings sustainability	A summary score based on firm fundamentals that informs about the sustainability of earning	Individual accounting	*Working Paper*	Penman and Zhang (2002)
2002	33		Market illiquidity	Average over the year of the daily ratio of the stock's absolute return to its dollar trading volume	Common microstructure	*Journal of Financial Markets*	Amihud (2002)
2003	34		GDP growth news	GDP growth news obtained from predictive regressions on lagged equity and fixed-income portfolios	Common macro	*Journal of Financial Economics*	Vassalou (2003)
2003	35		Market liquidity	Aggregated liquidity based on firm future excess stock return regressed on current signed excess return times trading volume	Common microstructure	*Journal of Political Economy*	Pastor and Stambaugh (2003)
2003			Idiosyncratic return volatility†	Residual variance obtained by regressing daily stock returns on market index return	Individual financial	*Journal of Financial Economics*	Ali, Hwang and Trombley (2003)
			Transaction costs†	Bid-ask spread, volume, etc.	Individual microstructure		
			Investor sophistication†	Number of analysts or institutional owners	Individual accounting		
2003		51	Shareholder rights	Shareholder rights as proxied by an index using 24 governance rules	Individual accounting	*Quarterly Journal of Economics*	Gompers, Ishii and Metrick (2003)

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2003		52	Excluded expenses	Excluded expenses in firm's earnings reports	Individual accounting	*Review of Accounting Studies*	Jeffrey, Lundholm and Soliman (2003)
2003		53	Growth in long-term net operating assets	Growth in long-term net operating assets	Individual accounting	*Accounting Review*	Fairfield, Whisenant and Yohn (2003)
2003		54	Order backlog	Order backlog divided by average total assets, transformed to a scaled-decile variable	Individual accounting	*Review of Accounting Studies*	Rajgopal, Shevlin and Venkatachalam (2003)
2003		55	Return consistency	Consecutive returns with the same sign	Individual financial	*Journal of Behavioral Finance*	Watkins (2003)
2004	36		Idiosyncratic consumption	Cross-sectional consumption growth variance	Common macro	*Journal of Finance*	Jacobs and Wang (2004)
2004	37		Cash flow news	News about future market cash flow	Common financial	*American Economic Review*	Campbell and Vuolteenaho (2004)
	38		Discount rate news	News about future market discount rate	Common financial		
2004			Market return†	Equity index return	Common financial	*Review of Financial Studies*	Vanden (2004)[g]
	39		Index option returns	Return on S&P 500 index option	Common financial		
2004	40		Default risk	Firm default likelihood using Merton's option pricing model	Common financial	*Journal of Finance*	Vassalou and Xing (2004)
2004	41		Real interest rate	Real interest rates extracted from a time-series model of bond yields and expected inflation	Common financial	*Journal of Finance*	Brennan, Wang and Xia (2004)
	42		Maximum Sharpe ratio portfolio	Maximum Sharpe ratio portfolio extracted from a time-series model of bond yields and expected inflation	Common financial		
2004	43		Return reversals at the style level	Zero-investment portfolios sorted based on past return performance at the style level	Common other	*Journal of Financial Economics*	Teo and Woo (2004)

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2004		56	Unexpected change in R&D	Unexpected change in firm research and expenditures	Individual accounting	*Journal of Finance*	Allan, Maxwell and Siddique (2004)
2004		57	52-week high	Nearness to the 52-week high price	Individual financial	*Journal of Finance*	George and Hwang (2004)
2004		58	Analysts' recommendations	Consensus analysts' recommendations from sell-side firms	Individual accounting	*Journal of Finance*	Jegadeesh, Kim, Krische and Lee (2004)
2004		59	Put-call parity	Violations of put-call parity	Individual financial	*Journal of Financial Economics*	Ofek, Richardson and Whitelaw (2004)
2004		60	Abnormal capital investment	Past year capital expenditures scaled by average capital expenditures for previous three years	Individual accounting	*Journal of Financial and Quantitative Analysis*	Titman, Wei and Xie (2004)
2005	44		Long-horizon consumption growth	Three-year consumption growth rate	Common macro	*Journal of Political Economy*	Parker and Julliard (2005)
2005	45		Long-run consumption	Cash flow risk measured by cointegration residual with aggregate consumption	Common macro	*Journal of Finance*	Bansal, Dittmar and Lundblad (2005)
2005	46		Housing price ratio	Ratio of housing to human wealth	Common financial	*Journal of Finance*	Lustig and Nieuwerburgh (2005)
2005		61	External corporate governance	Proxies for corporate control	Individual accounting	*Journal of Finance*	Cremers and Nair (2005)
		61	Internal corporate governance	Proxies for share-holder activism	Individual accounting		
2005			Market return†	Equity index return	Common financial	*Journal of Financial Economics*	Acharya and Pedersen (2005)[h]
	47		Market liquidity*	Value-weighted individual stock illiquidity as defined in Amihud (2002)	Common microstructure		
		63	Individual stock liquidity	Individual stock illiquidity as defined in Amihud (2002)	Individual microstructure		
2005		64	Price delay	Delay in a stock price's response to information	Individual microstructure	*Review of Financial Studies*	Hou and Moskowitz (2005)

…continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2005		65	Heterogeneous beliefs	Factors constructed from disagreement among analysts about expected short- and long-term earnings	Individual financial	*Review of Financial Studies*	Anderson, Ghysels and Juergens (2005)
2005		66	Short-sale constraints	Short-sale constraint proxied by Institutional ownership	Individual microstructure	*Journal of Financial Economics*	Nagel (2005)
2005		67	Short-sale constraints	Short-sale constraint proxied by short interest and institutional ownership	Individual microstructure	*Journal of Financial Economics*	Asquith, Pathak and Ritter (2005)
2005		68	Patent citation	Change of patent citation impact deflated by average total assets	Individual other	*Journal of Accounting, Auditing & Finance*	Gu (2005)
2005		69	Information uncertainty	Information uncertainty proxied by firm age, return volatility, trading volume or cash flow duration	Individual financial	*Review of Accounting Studies*	Jiang, Lee and Zhang (2005)
2005		70	Adjusted R&D	Adjusted R&D that incorporates capitalization and amortization	Individual accounting	*Working Paper*	Lev, Nissim and Thomas (2005)
2005		71	R&D reporting biases	R&D reporting biases proxied by the difference between R&D growth and earnings growth	Individual accounting	*Contemporary Accounting Research*	Lev, Sarath and Sougiannis (2005)
2005		72	Growth index	A combined index constructed based on earnings, cash flows, earnings stability, growth stability and intensity of R&D, capital expenditure and advertising	Individual accounting	Review of Accounting Studies	Mohanram (2005)
2006			Market return†	Equity index return and its square	Common financial	*Review of Financial Studies*	Vanden (2006)^{i}
			Index option return†	Index option return and its square	Common financial		
			Interaction between index and option return‡	Product of market and option returns	Common financial		

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2006	48		Financing frictions	Default premium	Common financial	*Review of Financial Studies*	Gomes, Yaron and Zhang (2006)
2006	49		Investment growth by households*	Household investment growth	Common macro	*Journal of Business*	Li, Vassalou and Xing (2006)
	50		Investment growth by nonfarm nonfinancial corporate firms	Nonfarm nonfinancial corporate firms investment growth	Common macro		
	51		Investment growth by nonfarm noncorporate business	Nonfarm noncorporate business investment growth	Common macro		
	52		Investment growth by financial firms	Financial firms investment growth	Common macro		
2006			Third to tenth power of market return‡	Third to tenth power of market return	Common financial	*Journal of Business*	Chung, Johnson and Schill (2006)[j]
2006		73	Financial constraints	Constraint index estimated from a firm's investment Euler equation	Individual financial	*Review of Financial Studies*	Whited and Wu (2006)
2006	53		Downside risk	Correlation with index return conditional on index return being below a threshold value	Common financial	*Review of Financial Studies*	Ang, Chen and Xing (2006)
2006	54		Systematic volatility	Aggregate volatility relative to Fama and French (1992) three-factor model	Common financial	*Journal of Finance*	Ang, Hodrick, Xing and Zhang (2006)
		74	Idiosyncratic volatility	Idiosyncratic volatility relative to Fama and French (1992) three-factor model	Individual financial		
2006	55		Investor sentiment	Composite sentiment index based on various sentiment measures	Common behavioral	*Journal of Finance*	Baker and Wurgler (2006)
2006	56		Retail investor sentiment	Systematic retail trading based on transaction data	Common behavioral	*Journal of Finance*	Kumar and Lee (2006)
2006	57		Durable and nondurable consumption growth	Durable and nondurable consumption growth	Common macro	*Journal of Finance*	Yogo (2006)
2006			Market return†	Equity index return	Common financial	*Journal of Finance*	Lo and Wang (2006)

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
	58		Trading volume	Return on a hedge portfolio constructed using trading volume and market returns	Common microstructure		
2006	59		Liquidity	Market-wide liquidity constructed first by decomposing firm-level liquidity into variable and fixed price effects then averaging the variable component	Common microstructure	*Journal of Financial Economics*	Sadka (2006)
2006	60		Earnings	Return on a zero-investment portfolio long in stocks with high earnings surprises and short in stocks with low earnings surprises	Common accounting	*Journal of Financial Economics*	Chordia and Shivakumar (2006)
2006	61		Liquidity	Turnover-adjusted number of days with zero trading over the prior 12 months	Common microstructure	*Journal of Financial Economics*	Liu (2006)
2006		75	Capital investment	Capital expenditure growth	Individual accounting	*Journal of Finance*	Anderson and Garcia-Feijoo (2006)
2006		76	Industry concentration	Industry concentration as proxied by the Herfindahl index	Individual accounting	*Journal of Finance*	Hou and Robinson (2006)
2006		77	Environment indicator*	A composite index measuring a firm's environmental responsibility	Individual other	*Financial Management*	Brammer, Brooks and Pavelin (2006)
		78	Employment indicator*	A composite index measuring employee responsibility	Individual other		
		79	Community indicator*	A composite index measuring community responsiveness	Individual other		
2006		80	Intangible information	Residuals from cross-sectional regression of firm returns on fundamental growth measures	Individual accounting	*Journal of Finance*	Daniel and Titman (2006)

...continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2006		81	Profitability	Expected earnings growth	Individual accounting	*Journal of Financial Economics*	Fama and French (2006)
		82	Investment*	Expected growth in book equity	Individual accounting		
			Book-to-market†	Book value of equity plus deferred taxes to market value of equity	Individual accounting		
2006		83	Net financing	Net amount of cash flow received from external financing	Individual accounting	*Journal of Accounting and Economics*	Bradshaw, Richardson and Sloan (2006)
2006		84	Forecasted earnings per share	Analysts' forecasted earnings per share	Individual accounting	*Working Paper*	Cen, Wei and Zhang (2006)
2006		85	Pension plan funding	Pension plan funding status calculated as the difference between the fair value of plan assets and the projected benefit obligation, divided by market capitalization	Individual accounting	*Journal of Finance*	Franzoni and Marin (2006)
2006		86	Acceleration	Firm's ranking on change in six-month momentum relative to the cross-section of other firms	Individual financial	*Working Paper*	Gettleman and Marks (2006)
2006		87	Unexpected earnings' autocorrelations	Standardized unexpected earnings' autocorrelations via the sign of the most recent earnings realization	Individual accounting	*Journal of Accounting Research*	Narayanamoorthy (2006)
2007	62		Payout yield	Return on a zero-investment portfolio long in high-yield stocks and short in low-yield stocks	Common accounting	*Journal of Finance*	Boudoukh, Michaely, Richardson and Roberts (2007)
2007	63		Productivity	Productivity level as in King and Rebelo (2000)	Common macro	*Journal of Financial Economics*	Balvers and Huang (2007)
	64		Capital stock	Quarterly capital stock interpolated from annual data	Common macro		
2007	65		Fourth-quarter to fourth-quarter consumption growth	Fourth-quarter to fourth-quarter consumption growth rate	Common macro	*Journal of Finance*	Jagannathan and Wang (2007)

...continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2007		88	Credit rating	S&P firm credit rating	Individual financial	*Journal of Finance*	Avramov, Chordia, Jostova and Philipov (2007)
2007		89	Trader composition	Fraction of total trading volume of a stock from institutional trading	Individual microstructure	*Working Paper*	Shu (2007)
2007		90	Change in order backlog	Change in order backlog	Individual accounting	*Seoul Journal of Business*	Baik and Ahn (2007)
2007		91	Firm productivity	Firm productivity measured by returns on invested capital	Individual accounting	*Working Paper*	Brown and Rowe (2007)
2007		92	Insider forecasts of firm volatility	Future firm volatility obtained from executive stock options	Individual financial	*Working Paper*	James, Fodor and Peterson (2007)
2007		93	Ticker symbol	Creativity in stocks' ticker symbols	Individual other	*Quarterly Review of Economics & Finance*	Head, Smith and Wilson (2007)
2007	66		Earnings cyclicality	Sensitivity of earnings to changes in aggregate total factor productivity	Common macro	*Working Paper*	Gourio (2007)
2008	67		Market volatility innovation	Difference in monthly average of squared daily return differences	Common financial	*Review of Financial Studies*	Kumar, Sorescu, Boehme and Danielsen (2008)
		94	Firm age	Firm's public listing age	Individual accounting		
			Market return†	Equity index return	Common financial		
		95	Interaction between market volatility and firm age	Product of market volatility and firm age	Individual accounting		
2008	68		Short-run market volatility	High frequency volatility extracted from a time-series model of market returns	Common financial	*Journal of Finance*	Adrian and Rosenberg (2008)
	69		Long-run market volatility	Low frequency volatility extracted from a time-series model of market returns	Common financial		

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2008	70		Investment growth	Return on a zero-investment portfolio long in low investment growth firms and short in high investment growth firms	Common financial	*Review of Financial Studies*	Xing (2008)
2008	71		Mean consumption growth	Across-state mean consumption growth rate	Common macro	*Review of Financial Studies*	Korniotis (2008)
	72		Variance of consumption growth*	Across-state consumption growth variance	Common macro		
	73		Mean habit growth	Across-state mean habit growth rate	Common macro		
	74		Variance of habit growth	Across-state habit growth variance	Common macro		
2008	75		Liquidity	Systematic liquidity extracted from eight empirical liquidity measures	Common microstructure	*Journal of Financial Economics*	Korajczyk and Sadka (2008)
2008		96	Country-level idiosyncratic volatility	Weighted average of variances and auto-covariances of firm-level idiosyncratic return shocks	Individual financial	*Review of Financial Studies*	Guo and Savickas (2008)
2008		97	Distress	Distressed firm failure probability estimated based on a dynamic logit model	Individual financial	*Journal of Finance*	Campbell, Hilscher and Szilagyi (2008)
2008		98	Shareholder advantage	Benefits from renegotiation upon default	Individual accounting	*Review of Financial Studies*	Garlappi, Shu and Yan (2008)
		99	Interaction between shareholder advantage and implied market value of assets	Implied market value of assets provided by Moody's KMV	Individual accounting		
2008		100	Asset growth	Year-on-year percentage change in total assets	Individual accounting	*Journal of Finance*	Cooper, Gulen and Schill (2008)

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2008		101	Share issuance	Annual share issuance based on adjusted shares	Individual accounting	*Journal of Finance*	Pontiff and Woodgate (2008)
2008			Earnings announcement return†	Earnings announcement return capturing the market reaction to unexpected information contained in the firm's earnings release	Individual financial	*Working Paper*	Brandt, Kishore, Santa-Clara and Venkatachalam (2008)
2008		102	Firm economic links	Economic links proxied by return of a portfolio of its major customers	Individual financial	*Journal of Finance*	Cohen and Frazzini (2008)
2008		103	Sin stock	Stocks in the industry of adult services, alcohol, defense, gaming, medical and tobacco	Individual other	*Financial Analyst Journal*	Frank, Ma and Oliphant (2008)
2008		104	Goodwill impairment	Buyers' overpriced shares at acquisition	Individual accounting	*Accounting Review*	Gu and Lev (2008)
2008		105	Information in order backlog	Changes in order backlog on future profitability	Individual accounting	*Working Paper*	Gu, Wang and Ye (2008)
2008		106	Investor recognition	Investor recognition proxied by the change in the breadth of institutional ownership	Individual other	*Review of Accounting Studies*	Lehavy and Sloan (2008)
2008		107	DuPont analysis	Sales over net operating assets in DuPont analysis	Individual accounting	*Accounting Review*	Soliman (2008)
2008		108	Small trades	Volume arising from small trades	Individual microstructure	*Review of Financial Studies*	Hvidkjaer (2008)
2008	76		Idiosyncratic component of S&P 500 return	Residual of the linear projection of the S&P 500 return onto the CRSP value weighted index return	Common financial	*Working Paper*	Brennan and Li (2008)
2009	77		Cash flow covariance with aggregate consumption	Cash flow covariance with aggregate consumption	Common macro	*Journal of Finance*	Da (2009)
	78		Cash flow duration	Cash flow duration sensitivity to aggregate consumption	Common macro		
2009			Financial constraints	THEORY	Common financial/macro	*Journal of Finance*	Livdan, Sapriza and Zhang (2009)

…continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2009	79		Long-run stockholder consumption growth	Aggregated microlevel stockholder consumption	Common macro	*Journal of Finance*	Malloy, Moskowitz and Vissing-Jorgensen (2009)
2009	80		Takeover likelihood	Estimated via a logit model of regressing ex-post acquisition indicator on various firm- and industry-level accounting variables	Common financial	*Review of Financial Studies*	Cremers, Nair and John (2009)
2009	81		Illiquidity	Estimated using structural formula in line with Kyle's (1985) lambda	Common microstructure	*Review of Financial Studies*	Chordia, Huh and Subrahmanyam (2009)
2009	82		Cash flow	Aggregate earnings based on revisions to analyst earnings forecasts	Common accounting	*Journal of Financial Economics*	Da and Warachka (2009)
2009	83		Investors' beliefs*	Belief extracted from a two-state regime-switching model of aggregate market return and aggregate output	Common other	*Review of Financial Studies*	Ozoguz (2008)
	84		Investors' uncertainty	Uncertainty extracted from a two-state regime-switching model of aggregate market return and aggregate output	Common other		
2009		109	Media coverage	Firm mass media coverage	Individual behavioral	*Journal of Finance*	Fang and Peress (2009)
2009		110	Financial distress	Credit rating downgrades	Individual accounting	*Journal of Financial Economics*	Avramov, Chordia, Jostova and Philipov (2009)
2009		111	Idiosyncratic volatility	Conditional expected idiosyncratic volatility estimated from a GARCH model	Individual accounting	*Journal of Financial Economics*	Fu (2009)
2009		112	Debt capacity	Firm tangibility as in Almeida and Campello (2007)	Individual accounting	*Journal of Finance*	Hahn and Lee (2009)

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2009		113	Realized-implied volatility spread	Difference between past realized volatility and the average of call and put implied volatility	Individual financial	*Management Science*	Bali and Hovakimian (2009)
		114	Call-put implied volatility spread	Difference between call and put implied volatility	Individual financial		
2009		115	Productivity of cash	Net present value of all the firm's present and future projects generated per dollar of cash holdings	Individual accounting	*Working Paper*	Chandrashekar and Rao (2009)
2009		116	Advertising	Change in expenditures on advertising	Individual accounting	*Working Paper*	Chemmanur and Yan (2009)
2009		117	Analyst forecasts optimism	Relative optimism and pessimism proxied by the difference between long-term and short-term analyst forecast of earnings growth	Individual financial	*Journal of Financial Markets*	Da and Warachka (2009)
2009		118	Information revelation	Monthly estimate of the daily correlation between absolute returns and dollar volume	Individual microstructure	*Working Paper*	Gokcen (2009)
2009		119	Earnings volatility	Earnings volatility	Individual accounting	*Working Paper*	Gow and Taylor (2009)
2009		120	Cash flow volatility	Rolling standard deviation of the standardized cashflow over the past sixteen quarters	Individual accounting	*Journal of Empirical Finance*	Huang (2009)
2009		121	Local unemployment	Relative state unemployment	Individual other	*Working Paper*	Korniotis and Kumar (2009)
		122	Local housing collateral	State-level housing collateral	Individual other		
2009		123	Efficiency score	Firm efficiency/inefficiency identified from the residual of the projection of firm market-to-book ratio onto various firm financial and accounting variables	Individual financial	*Journal of Financial and Quantitative Analysis*	Nguyen and Swanson (2009)

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2009		124	Order imbalance	Difference between buyer- and seller-initiated trades	Individual microstructure	*Review of Financial Studies*	Barber, Odean and Zhu (2009)
2010	85		Market volatility and jumps	Estimated based on S&P index option returns	Common financial	*Working Paper*	Cremers, Halling and Weinbaum (2010)
2010	86		Market mispricing	Zero-investment portfolio constructed from repurchasing and issuing firms	Common behavioral	*Review of Financial Studies*	Hirshleifer and Jiang (2010)
2010		125	Idiosyncratic skewness	Skewness forecasted using firm level predictive variables	Individual financial	*Review of Financial Studies*	Boyer, Mitton and Vorkink (2010)
2010		126	Political campaign contributions	Firm contributions to US political campaigns	Individual other	*Journal of Finance*	Cooper, Gulen and Ovtchinnikov (2010)
2010		127	Real estate holdings	Real estate to total property, plant and equipment	Individual accounting	*Review of Financial Studies*	Tuzel (2010)
2010		128	Realized skewness	Realized skewness obtained from high-frequency intraday prices	Individual financial	*Working Paper*	Amaya, Christoffersen, Jacobs and Vasquez (2011)
		129	Realized kurtosis	Realized kurtosis obtained from high-frequency intraday prices	Individual financial		
2010		130	Excess multiple	Excess multiple calculated as the difference between the accounting multiple and the warranted multiple obtained by regressing the cross-section of firm multiples on accounting variables	Individual accounting	*Journal of Accounting, Auditing & Finance*	An, Bhojraj and Ng (2010)
2010		131	Firm information quality	Firm information quality proxied by analyst forecasts, idiosyncratic volatility and standard errors of beta estimates	Individual financial/accounting	*Working Paper*	Armstrong, Banerjee and Corona (2010)
2010		132	Long-run idiosyncratic volatility	Long-run idiosyncratic volatility filtered from idiosyncratic volatility using HP filters	Individual financial	*Working Paper*	Cao and Xu (2010)

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2010	87		Private information	Return on a zero-investment portfolio long in high PIN stocks and short in low PIN stocks; PIN (private information) is the probability of information- based trade	Common microstructure	*Journal of Financial and Quantitative Analysis*	David, Hvidkjaer and O'Hara (2010)
2010		133	Intra-industry return reversals	Intra-industry return reversals captured by the return difference between loser stocks and winners stocks based on relative monthly performance within the industry	Individual financial	*Working Paper*	Hameed, Huang and Mian (2010)
2010		134	Related industry returns	Stock returns from economically related supplier and customer industries	Individual financial	*Journal of Finance*	Menzly and Ozbas (2010)
2010		135	Earnings distributed to equity holders	Earnings distributed to equity holders	Individual accounting	*Review of Accounting & Finance*	Papanastasopoulos, Thomakos and Wang (2010)
		136	Net cash distributed to equity holders	Dividends minus stock issues	Individual accounting		
2010		137	Excess cash	Most recently available ratio of cash to total assets	Individual accounting	*Financial Management*	Simutin (2010)
2010		138	Extreme downside risk	Extreme downside risk proxied by the left tail index in the classical generalized extreme value distribution	Individual financial	*Journal of Banking and Finance*	Huang, Liu, Rhee and Wu (2010)
2010		139	Volatility smirk	Steepness in individual option volatility smirk	Individual financial	*Journal of Financial and Quantitative Analysis*	Xing, Zhang and Zhao (2010)
2010			Exposure to financial distress costs	THEORY	Individual financial	*Journal of Financial Economics*	George and Hwang (2010)
2011	88		Rare disasters	Disaster index based on international political crises	Common financial	*Journal of Financial Economics*	Berkman, Jacobsen and Lee (2011)

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2011			Distress risk‡	Aggregate distress risk obtained by projecting future business failure growth rates on a set of basis assets	Common financial	*Journal of Financial Economics*	Kapadia (2011)[k]
2011			Momentum†	Factor-mimicking portfolios based on momentum of international equity returns	Common other	*Review of Financial Studies*	Hou, Karolyi and Kho (2011)
	89		Cash flow-to-price	Factor-mimicking portfolios based on cash flow-to-price of international equity returns	Common accounting		
2011		140	R&D investment	Firm's investment in research and development	Individual accounting	*Review of Financial Studies*	Li (2011)
			Financial constraints†	Kaplan and Zingales (1997) financial constraint index	Individual financial		
2011		141	Extreme stock returns	Portfolios sorted based on extreme past returns	Individual financial	*Journal of Financial Economics*	Bali, Cakici and Whitelaw (2011)
2011		142	Jumps in individual stock returns	Average jump size proxied by slope of option implied volatility smile	Individual financial	*Journal of Financial Economics*	Yan (2011)
2011		143	Intangibles	Employee satisfaction proxied by the list of "100 Best Companies to Work for in America"	Individual other	*Journal of Financial Economics*	Edmans (2011)
2011			Market return†	Equity index return	Common financial	*Working Paper*	Chen, Novy-Marx and Zhang (2011)
	90		Investment portfolio return	Difference between returns of portfolios with low and high investment-to-asset ratio	Common financial		
	91		Return-on-equity portfolio return	Difference between returns of portfolios with high and low return on equity	Common financial		
2011		144	Volatility of liquidity	Measured by the price impact of trade as in Amihud (2002)	Individual microstructure	*Working Paper*	Akbas, Armstrong and Petkova (2011)

...continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2011		145	Dispersion in beliefs	Revealed through active holdings of fund managers	Individual behavioral	*Working Paper*	Jiang and Sun (2011)
2011		146	Credit default swap spreads	Five-year spread less one-year spread	Individual financial	*Working Paper*	Han and Zhou (2011)
2011		147	Organizational capital	Directly measured using Selling, General and Administrative expenditures	Individual accounting	*Working Paper*	Eisfeldt and Papanikolaou (2011)
2011		148	Residual income	Firm residual income growth extracted from firm earnings growth	Individual accounting	*Review of Accounting Studies*	Balachandran and Mohanram (2011)
2011		149	Accrual volatility	Firm accrual volatility measured by the standard deviation of the ratio of accruals to sales	Individual accounting	*Working Paper*	Bandyopadhyay, Huang and Wirjanto (2011)
2011		150	Implied cost of capital	Implied cost of capital estimated using option contracts	Individual financial	*Working Paper*	Callen and Lyle (2011)
2011		151	Non-accounting information quality	Average delay with which non-accounting information is impounded into stock price	Individual financial	*Contemporary Accounting Research*	Callen, Khan and Lu (2011)
		152	Accounting information quality	Average delay with which accounting information is impounded into stock price	Individual financial		
2011		153	Labor unions	Labor force unionization measured by the percentage of employed workers in a firm's primary Census industry Classification industry covered by unions in collective bargaining with employers	Individual other	*Journal of Financial and Quantitative Analysis*	Chen, Kacperczyk and Ortiz-Molina (2011)

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2011		154	Overreaction to nonfundamental price changes	Overreaction to within-industry discount rate shocks as captured by decomposing the short-term reversal into across-industry return momentum, within-industry variation in expected returns, under-reaction to within-industry cash flow news and overreaction to within-industry discount rate news	Individual other	*Working Paper*	Da, Liu and Schaumburg (2011)
2011		155	Short interest	Short interest from short sellers	Individual financial	*Accounting Review*	Michael and Rees (2011)
2011		156	Percent total accrual	Firm accruals scaled by earnings	Individual accounting	*Accounting Review*	Hafzalla, Lundholm and Van Winkle (2007)
2011			Projected earnings accuracy‡	Skilled analysts identified by both past earnings forecasts accuracy and skills	Individual accounting	*Working Paper*	Hess, Kreutzmann and Pucker (2011)
2011		157	Firm productivity	Firm level total factor productivity estimated from firm value added, employment and capital	Individual accounting	*Working Paper*	Imrohoroglu and Tuzel (2011)
2011		158	Really dirty surplus	Really dirty surplus that happens when a firm issues or reacquires its own shares in a transaction that does not record the shares at fair market value	Individual accounting	*Accounting Review*	Landsman, Miller, Peasnell and Shu (2011)
2011		159	Earnings forecast	Earnings forecast based on firm fundamentals	Individual accounting	*Review of Accounting Studies*	Li (2011)
2011		160	Asset growth	Yearly percentage change in total balance sheet assets	Individual accounting	*Working Paper*	Nyberg and Poyry (2011)
2011		161	Real asset liquidity	Number of potential buyers for a firm's assets from within the industry	Individual microstructure	*Working Paper*	Ortiz-Molina and Phillips (2011)

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2011		162	Customer-base concentration	Annual change in customer-base concentration	Individual other	*Working Paper*	Patatoukas (2011)
2011		163	Tax expense surprises	Seasonally differenced quarterly tax expense	Individual accounting	*Journal of Accounting Research*	Thomas and Zhang (2011)
2011			Predicted earnings increase score†	Predicted earnings increase score based on financial statement information	Individual accounting	*Review of Accounting Studies*	Wahlen and Wieland (2011)
2011			Shareholder recovery	THEORY	Common financial	*Journal of Finance*	Garlappi and Yan (2011)
2011	92		Garbage growth	Realized annual garbage growth	Common macro	Journal of Finance	Savov (2011)
2012	93		Financial intermediary's wealth	Intermediary's marginal value of wealth proxied by shocks to leverage of securities broker-dealers	Common financial	*Journal of Finance*	Adrian, Etula and Muir (2012)
2012	94		Stochastic volatility*	Estimated from a heteroscedastic VAR based on market and macro variables	Common financial	*Working Paper*	Campbell, Giglio, Polk and Turley (2012)
2012	95		Average variance of equity returns	Decomposition of market variance into an average correlation component and an average variance component	Common financial	*Review of Financial Studies*	Chen and Petkova (2012)
2012	96		Income growth for goods producing industries	Income growth for goods producing industries	Common macro	*Journal of Finance*	Eiling (2012)
	97		Income growth for manufacturing industries	Income growth for manufacturing industries	Common macro		
	98		Income growth for distributive industries	Income growth for distributive industries	Common macro		
	99		Income growth for service industries*	Income growth for service industries	Common macro		
	100		Income growth for government*	Income growth for government	Common macro		

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2012	101		Consumption volatility	Filtered consumption growth volatility from a Markov regime-switching model based on historical consumption data	Common macro	*Journal of Finance*	Boguth and Kuehn (2012)
2012	102		Market skewness	Higher moments of market returns estimated from daily index options	Common financial	*Journal of Financial Economics*	Chang, Christoffersen and Jacobs (2012)
2012	103		Learning*	Learning estimated from an investor's optimization problem under Knightian uncertainty	Common financial	*Working Paper*	Viale, Garcia-Feijoo and Giannetti (2011)
	104		Knightian uncertainty	Knightian uncertainty estimated from an investor's optimization problem under Knightian uncertainty	Common financial		
2012	105		Market uncertainty	Proxied by variance risk premium	Common financial	*Working Paper*	Bali and Zhou (2012)
2012			Labor income‡	Labor income at the census division level	Common macro	*Working Paper*	Gomez, Priestley and Zapatero (2012)^{l}
2012		164	Product price change	Cumulative product price changes since an industry enters the producer price index program	Individual financial	*Working Paper*	Van Binsbergen (2012)
2012	106		Future growth in the opportunity cost of money	Opportunity cost of money as proxied by 3-month Treasury bill rate or effective Federal Funds rate	Common macro	*Working Paper*	Lioui and Maio (2012)
2012			Inter-cohort consumption differences	THEORY	Common macro	*Journal of Financial Economics*	Garleanu, Kogan and Panageas (2012)
2012	107		Market-wide liquidity	Proxied by "noise" in Treasury prices	Common microstructure	*Working Paper*	Hu, Pan and Wang (2012)
2012		165	Stock skewness	Ex ante stock risk-neutral skewness implied by option prices	Individual financial	*Journal of Finance*	Conrad, Dittmar and Ghysels (2012)

...continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2012		166	Expected return uncertainty	Proxied by the volatility of option-implied volatility	Individual financial	*Working Paper*	Baltussen, Van Bekkum and Van der Grient (2012)
2012		167	Information intensity	Proxied by monthly frequency of current report filings	Individual microstructure	*Working Paper*	Zhao (2012)
2012		168	Credit risk premia	Market implied credit risk premia based on the term structure of CDS spreads	Individual financial	*Working Paper*	Friewald, Wagner and Zechner (2012)
2012		169	Geographic dispersion	Number of states in which a firm has business operations	Individual other	*Journal of Financial Economics*	Garcia and Norli (2012)
2012		170	Political geography	Political proximity measured by political alignment index of each state's leading politicians with the ruling presidential party	Individual other	*Journal of Financial Economics*	Kim, Pantzalis and Park (2012)
2012		171	Option to stock volume ratio	Option volume divided by stock volume	Individual microstructure	*Journal of Financial Economics*	Johnson and So (2012)
2012		172	Cash holdings	Firm cash holdings	Individual accounting	*Journal of Financial Economics*	Palazzo (2012)
2012		173	Labor mobility	Labor mobility based on average occupational dispersion of employees in an industry	Individual accounting	*Working Paper*	Donangelo (2012)
2012		174	Debt covenant protection	Firm-level covenant index constructed based on 30 covenant categories	Individual accounting	*Working Paper*	Wang (2012)
2012		175	Stock-cash flow sensitivity	Stock-cash flow sensitivity estimated from a structural one-factor contingent-claim model	Individual financial	*Working Paper*	Chen and Strebulaev (2012)
2012	108		Jump beta	Discontinuous jump beta based on Todorov and Bollerslev (2010)	Common financial	*Working Paper*	Sophia Zhengzi Li (2012)

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2012			Long-run consumption growth‡	Long-run consumption growth rate identified from the risk-free rate and market price-dividend ratio based on Bansal and Yaron (2005)'s long-run risk model	Common macro	*Journal of Financial Economics*	Ferson, Nallareddy and Xie (2012)^{m}
			Short-run consumption growth‡	Short-run consumption growth rate identified from the risk-free rate and market price-dividend ratio based on Bansal and Yaron (2005)'s long-run risk model	Common macro		
			Consumption growth volatility‡	Consumption growth volatility shocks identified from the risk-free rate and market price-dividend ratio based on Bansal and Yaron (2005)'s long-run risk model	Common macro		
2012		176	Change in call implied volatility	Change in call implied volatility	Individual financial	*Working Paper*	Ang, Bali and Cakici (2012)
		177	Change in put implied volatility	Change in put implied volatility	Individual financial		
2012		178	Firm hiring rate	Firm hiring rate measured by the change in the number of employees over the average number of employees	Individual other	*Working Paper*	Bazdresch, Belo and Lin (2012)
2012		179	Information processing complexity	Past return for paired pseudo-conglomerates	Individual financial	*Journal of Financial Economics*	Cohen and Lou (2012)
2012		180	Opportunistic buy	Prior month buy indicator for opportunistic traders who do not trade routinely	Individual microstructure	*Journal of Finance*	Cohen, Malloy and Pomorski (2012)
		181	Opportunistic sell	Prior month sell indicator for opportunistic traders who do not trade routinely	Individual microstructure		

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2012		182	Innovative efficiency	Patents/citations scaled by research and development expenditures	Individual other	*Journal of Financial Economics*	Hirshleifer, Hsu and Li (2012)
2012		183	Abnormal operating cash flows	Abnormal operating cash flows	Individual accounting	*Working Paper*	Li (2012)
		184	Abnormal production costs	Abnormal production costs	Individual accounting		
2012		185	Deferred revenues	Changes in the current deferred revenue liability	Individual accounting	*Contemporary Accounting Research*	Prakash and Sinha (2012)
2012		186	Earnings conference calls	Sentiment of conference call wording	Individual other	*Journal of Banking and Finance*	Price, Doran, Peterson and Bliss (2012)
2012		187	Earnings forecast optimism	Difference between characteristic forecasts and analyst forecasts	Individual accounting	*Working Paper*	So (2012)
2012	109		Commodity index	Open interest-weighted total index that aggregates 33 commodities	Common financial	*Working Paper*	Boons, Roon and Szymanowska (2012)
2012		188	Time-series momentum	Time-series momentum strategy based on autocorrelations of scaled returns	Individual financial	*Journal of Financial Economics*	Moskowitz, Ooi and Pedersen (2012)
2012		189	Carry	Expected return minus expected price appreciation	Individual financial	*Working Paper*	Koijen, Moskowitz, Pedersen and Vrugt (2012)
2012		190	Expected return proxy	Logistic transformation of the fit (R^2) from a regression of returns on past prices	Individual financial	*Journal of Financial Economics*	Burlacu, Fontaine, Jimenez-Garces and Seasholes (2012)
2012		191	Fraud probability	Probability of manipulation based on accounting variables	Individual accounting	*Financial Analysts Journal*	Beneish, Lee and Nichols (2013)
2012		192	Buy orders	Sensitivity of price changes to sell orders	Individual microstructure	*Working Paper*	Brennan, Chordia, Subrahmanyam and Tong (2012)
		193	Sell orders	Sensitivity of price changes to buy orders	Individual microstructure		

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2013	110		Expected dividend level	Expected dividend level based on a macro time-series model	Common financial	*Working Paper*	Doskov, Pekkala and Ribeiro (2013)
	111		Expected dividend growth	Expected dividend growth based on a macro time-series model	Common financial		
2013		194	Firm's ability to innovate	Rolling firm-by-firm regressions of firm-level sales growth on lagged R&D	Individual accounting	*Review of Financial Studies*	Cohen, Diether and Malloy (2013)
2013		195	Board centrality	Board centrality measured by four basic dimensions of well-connectedness	Individual other	*Journal of Accounting and Economics*	Larcker, So and Wang (2013)
2013		196	Gross profitability	Gross profits to assets	Individual accounting	*Journal of Financial Economics*	Novy-Marx (2013)
2013		197	Betting-against-beta	Long leveraged low-beta assets and short high-beta assets	Individual financial	*Working Paper*	Frazzini and Pedersen (2013)
2013		198	Secured debt	Proportion of secured to total debt	Individual accounting	*Working Paper*	Valta (2013)
		199	Convertible debt	Proportion of convertible to total debt	Individual accounting		
		200	Convertible debt indicator	Dummy variable indicating whether a firm has convertible debt outstanding	Individual accounting		
2013	112		Cross-sectional pricing inefficiency	Pricing inefficiency proxied by returns to simulated trading strategies that capture momentum, profitability, value, earnings and reversal	Common microstructure	*Working Paper*	Akbas, Armstrong, Sorescu and Subrahmanyam (2013)
2013		201	Attenuated returns	Composite trading strategy returns where the weights are based on averaging percentile rank scores of various characteristics for each stock on portfolios	Individual financial	*Working Paper*	Chordia, Subrahmanyam and Tong (2013)

. . . continued

Year	#	#	Factor	Formation	Type	Journal	Short reference
2013		202	Bad private information	Decomposing the PIN measure of Easley, Hvidkjaer and O'Hara (2002) into two elements that reflect informed trading on good news and bad news	Individual microstructure	*Working Paper*	Brennan, Huh and Subrahmanyam (2013)
2013	113		Trend signal	Return on a zero-investment portfolio long in past winners and short in past losers based on short-term, intermediate-term and long-term stock price trends	Common other	*Working Paper*	Han and Zhou (2013)

Notes to Table

This table contains a summary of risk factors that explain the cross-section of expected returns. The column "Indi.(#)" ("Common(#)") reports the cumulative number of empirical factors that are classified as individual (common) risk factors.

*: insignificant; †: duplicated; ‡: missing p-value.

a: No p-values reported for their factors constructed from principal component analysis.

b: Fama and French (1992) create zero-investment portfolios to test size and book-to-market effects. This is different from the testing approach in Banz (1981). We therefore count Fama and French (1992)'s test on size effect as a separate one.

c: No p-values reported for their high order equity index return factors.

d: No p-values reported for their eight risk factors that explain international equity returns.

e: No p-values reported for his high order return factors.

f: No p-values reported for their five hedge fund style return factors.

g: Vanden (2004) reports a t-statistic for each Fama-French 25 size and book-to-market sorted stock portfolios. We average these 25 t-statistics.

h: Acharya and Pedersen (2005) consider the illiquidity measure in Amihud (2002). This is different from the liquidity measure in Pastor and Stambaugh (2003). We therefore count their factor as a separate one.

i: No p-values reported for the interactions between market return and option returns.

j: No p-values reported for their co-moment betas.

k: No p-values reported for his distress tracking factor.

l: Gomez, Priestley and Zapatero (2012) study census division level labor income. However, most of the division level labor income have a non-significant t-statistic. We do not count their factors.

m: No p-values reported for their factors estimated from the long-run risk model.

5

Multiple Testing in Financial Economics

5.1 A Multiple Testing Framework

5.1.1 The Null Hypothesis

Assume that researchers have carried out M tests. The corresponding test statistics are given by $\mathbf{y} = (y_1, y_2, \ldots, y_M)$, with each y_i representing the test statistic or vector of test statistics for the i-th experiment. We want to perform M tests of hypotheses:

$$H_{0i}: \quad y_i \sim f_{0i},$$

$$H_{1i}: \quad y_i \sim f_{1i}.$$

Here f_{0i} and f_{1i} denote the null and alternative distribution, respectively. They often involve unknown parameters. Notice that at this level of generality, multiple testing amounts to considering this set of M hypotheses as a whole. We do not require information on the interrelationships among the f_{0i}'s or f_{1i}'s to design a specific testing method. For example, well-known procedures such as Bonferroni and Benjamini and Hochberg (1995)'s method can be applied to the collection of p-values of individual tests to control for family-wise error rate (FWER) and false-discovery rate (FDR), respectively.

In typical applications of multiple testing in economics, however, we do know the connections among the collection of null and alternative hypotheses and incorporating such information into the test procedure can improve the performance of the test. For instance, most studies that propose investment strategies based on the cross-section of stocks test the hypothesis that the mean strategy return is zero. These studies therefore have a common null hypothesis. Moreover, based on the primitive assumption of economic scarcity, it is reasonable to assume that more profitable strategies are less likely to exist. More specifically, among the truly profitable strategies (i.e., the alternative hypothesis is true), the number of strategies that achieve a certain level of mean return is declining in the level of the mean. Effectively, this imposes a monotonicity restriction on the density function of the average returns for profitable strategies. Both the common null observation and the monotonicity restriction under alternative hypotheses can be important to the design of a testing procedure as they help us better understand the composition of test statistics under both the null and the alternative hypotheses.

For the above reasons, we propose a model to facilitate the economic applications of multiple testing methods. We first parameterize f_{0i} and f_{1i} as a constant hypothesized value or vector μ_i. We assume that each μ_i is an independent draw from the following mixture distribution:

$$\mu_i \sim p_0 I_{\{\mu=0\}} + (1-p_0)F(\lambda),$$

where $I_{\{\mu=0\}}$ is the distribution that has a point mass at zero and $F(\lambda)$ is a family of distributions that are parameterized by vector λ.[1] This mixture distribution assumption is the key component for Bayesian hypothesis testing[2] and captures the dichotomy between the null and alternative hypothesis in a simple manner. From

[1]Without loss of generality, we assume that $\mu_i = 0$ is the common null of all hypotheses.

[2]See Meng and Dempster (1987), Scott and Berger (2006) and Whittemore (2007) for the Bayesian approach on multiple hypothesis testing.

the perspective of hierarchical modeling, the mixture distribution dictates that μ_i is generated in two steps. First, with probability p_0, μ_i assumes a value of zero and thus belongs to the null distribution. If μ_i does not equal zero in the first step, it is subsequently drawn from the parametric distribution $F(\lambda)$.

The parametric distribution $F(\lambda)$ is common across all tests and describes the distribution of μ_i when $\mu_i \neq 0$. Economic intuition often suggests distributional properties that $F(\lambda)$ should assume. For instance, when smaller values of μ_i are more plausible than larger ones due to economic scarcity, a single-parameter exponential family may be adequate to describe $F(\lambda)$. To allow for a more flexible shape and to separate the mean from the variance, a two-parameter Gamma distribution is an obvious extension of the exponential family. Similar to most econometric specifications, the bias and variance tradeoff applies and it is not always better to assume a more complicated model.

5.1.2 Decomposing Test Statistics

Given a population mean μ_i for the i-th hypothesis, we need a specification on the shock process to generate the associated test statistic. As shown in Lin (2005), all the commonly used statistics can be written or can be approximately by the statistics of the following form:

$$G_i = U_i^T V_i^{-1} U_i,\ i = 1, \ldots, M,$$

where

$$U_i = \sum_{k=1}^{n} U_{ik},$$

and

$$V_i = \sum_{k=1}^{n} U_{ik} U_{ik}^T.$$

Under the null hypothesis that $\mu_i = 0$, U_i approximately follows a normal distribution with mean zero and covariance matrix V_i in large samples, so that G_i approximately follows a χ^2 distribution with m_i degrees of freedom, where m_i is the dimension of U_i. In general, the U_i's are correlated and so are the G_i's. Although our most general framework is able to handle multi-dimensional U_i, we restrict our attention to the univariate case to ease exposition.

Guided by the general expression for test statistics above, we focus on the t statistic, which is the square root of the above χ^2 statistic and often used by economists. The t-test for testing the hypothesis $\mu_i = 0$ is given by

$$T_i = (\sum_{k=1}^{n} U_{ik}/n)/(\hat{\sigma}_i/\sqrt{n}),$$

where $\hat{\sigma}_i$ is an estimate of the standard deviation of $U_{ik}, k = 1, \ldots, n$. To model the cross-section of t-statistics and their correlations, we first model the cross-section of standard deviations. For each experiment i, suppose σ_i is independently drawn from the following distribution for standard deviations

$$\sigma_i \sim G(\xi),$$

where $G(\xi)$ is a family of distributions that are parameterized by vector ξ. Given the σ_i's, the scaled observation U_{ik}/σ_i has a mean of μ_i/σ_i and a standard deviation of one. We can now impose a correlation structure on both the cross-section and the time-series of U_{ik}/σ_i's to study its impact on multiple testing. The simplest structure is the equal correlation structure given by:

$$\begin{aligned} \mathrm{Corr}(U_{i,k}/\sigma_i, U_{j,k}/\sigma_j) &= \rho, \quad i \neq j, \\ \mathrm{Corr}(U_{i,k}/\sigma_i, U_{j,l}/\sigma_j) &= 0, \quad k \neq l. \end{aligned}$$

This correlation structure assumes zero correlation between variables that occur in different periods (both within and across experiments) and a correlation coefficient of ρ between any pair of contemporaneous variables. These modeling assumptions highlight the importance of correlation in the cross-section and make the model useful when the primary concern is about the impact of cross-sectional correlation on multiple testing. For a general correlation structure that extends beyond the equal correlation case, we use Φ to parameterize the correlation matrix. Φ contains the parameters in the correlation matrix that need to be estimated or calibrated.

More detailed forms of correlations can be imposed on the panel of $\{U_{ik}/\sigma_i, i = 1, \ldots, M, k = 1, \ldots, n\}$ to study their impacts on the formation of test statistics.[3] For instance, to accommodate time-series correlations within each experiment, we can use simple time-series models, e.g., an AR(1) process for $\{U_{ik}/\sigma_i, k = 1, \ldots, n\}$. After fixing the time-series model, we can either assume a constant autoregressive parameter across all experiments or a distribution from which the cross-section of autoregressive parameters are drawn. The latter assumption is more appropriate when there is a large degree of heterogeneity of shock persistence across experiments. Notationally, we collect the single autoregressive parameter or the parameters for the distribution of autoregressive parameters in Ψ.

An important issue is the separate identification of $F(\lambda)$ and $G(\tau)$. Suppose the standard deviation σ_i for $\{U_{ik}, k = 1, \ldots, n\}$ is known. Then the t-statistic is the sum of the scaled U_{ik}'s, that is,

$$\widetilde{T}_i = (\sum_{k=1}^{n} U_{ik}/\sigma_i)/\sqrt{n}.$$

Notice that each summand U_{ik}/σ_i has a mean of μ_i/σ_i and a variance of one. Assuming $\{U_{ik}\}_{k=1}^{n}$ are i.i.d., then $\widetilde{T}_i$ has a mean of $(\mu_i/\sigma_i)\sqrt{n}$ and a variance of one. Given that we only observe the individual t-statistics, the distribution F and G can only be identified through the distribution of $\{\mu_i/\sigma_i\}_{i=1}^{M}$. This observation prompts

[3] Viewing "i" as the individual index and "k" as the time index, we have a panel of variables.

us to model the distribution of the ratio μ_i/σ_i, as opposed to model the distribution of μ_i and σ_i separately. With a slight abuse of notation, we also use F to denote the distribution of μ_i/σ_i. In reality, σ_i is unknown so $\widetilde{T}_i$ differs from T_i by a factor of $\hat{\sigma}_i/\sigma_i$. However, this factor should be very close to one for large n as variance can be estimated with a high precision for time-series data with a relatively large number of observations.[4] Therefore, we use $\widetilde{T}_i$ to approximate T_i in our applications and focus on the estimation of the distribution of the signal-to-noise ratio μ_i/σ_i.

Finally, we assume that all shocks follow a multivariate normal distribution. When dependence is the concern, the multiple testing literature allows for a general dependence structure among test statistics and p-values.[5] This does not fit in our context as we need the exact distribution of shocks to generate simulated moments for test statistics. We choose to rely on the normal distribution for convenience and focus on dependence represented by Pearson correlations.[6]

In particular, suppose that $\{W_{ik}, k = 1, \ldots, n\}$ are the innovations for the time-series $\{U_{ik}/\sigma_i, k = 1, \ldots, n\}$ and are independent normal shocks with a variance of $1/(1-\rho_i^2)$, where ρ_i is the persistence for the process $\{U_{ik}/\sigma_i\}_{k=1}^n$. Cross-sectionally, $\{W_{ik}\}_{i=1}^M$ follows a multivariate normal distribution with a pre-imposed correlation structure (parameterized by Φ). Lastly, the vector of shocks $\{W_{ik}\}_{i=1}^M$ move independently in time (i.e., in k). These specifications, together with the specifications for the cross-section of means μ_i/σ_i, are sufficient for us to simulate the cross-section of t-statistics for any number of experiments M and time periods n. The simulated moments for the cross-section of t-statistics are then matched to the corresponding

[4]With n larger than one hundred, the Student's t distribution is indistinguishable from the normal distribution.

[5]See Benjamini and Yekutieli (2001) and Sarkar (2002) for their definitions of dependence and the reason why Benjamini and Hochberg (1995)'s procedure is robust against these forms of dependence.

[6]Note that any elliptical distribution, which includes the multivariate normal distribution, would work for our purpose.

moments of the observed t-statistics to estimate parameters of the model.

Essentially, our framework allows us to substitute normally distributed shocks for the unobservable U_{ik}'s. This substitution can be important for both the model estimation and the multiple testing adjustment based on the estimated model. While normality is an important and convenient case to consider, alternative distributional assumptions are also possible for other applications. For instance, certain investment strategy returns display higher order moment characteristics that can have a large impact on test statistics. When U_{ik}'s are observable, the literature on permutation tests suggest re-sampling techniques that are robust to shock specifications.[7] In this paper, we primarily focus on normal shocks to define both cross-sectional and time-series correlations and investigate how thus defined correlations affect multiple testing. Further investigation on alternative shock specifications is left for future research.

5.1.3 Publication Bias

Lastly, publication bias — not all research findings are published — is important for us to embed into our framework. Publication bias is a phenomenon that is likely endemic to all empirical inquires. In medical research, significant findings are more likely to be published and cited.[8] In economics, statistically significant results are more likely to be published.[9] That is, the editorial process makes it less likely that a paper testing an interesting hypothesis gets published if the results are insignificant.

Given this bias, published studies are unlikely to be representative of all studies that have been conducted. This undercoverage issue causes problems for multiple testing adjustments as all trials, including unpublished ones, need to be taken into

[7] See Ge et al. (2003) and Romano and Wolf (2005). A recent paper by Hsu, Hsu and Kuan (2010) applies the idea of permutation tests to evaluate mutual fund performances.

[8] The research on publication bias in medical research are voluminous. See Song et al. (2009) for a summary.

[9] See Fanelli (2012).

account. For instance, the commonly used Bonferroni adjustment works by multiplying a p-value by the number of tests. If publication bias is present, we are likely to underestimate the total number of tests. This leads to an inadequate adjustment and often causes the Type I error rates for multiple testing (e.g., FWER or FDR) to be above the pre-specified value.

To model publication bias, we assume that a study will be published if and only if its associated t-statistic exceeds a certain threshold value T_0. Our assumption is consistent with meta-analysis in the medical science literature.[10] Among the M studies that have been carried out, only M_0 of them are publicly available. Let the collection of t-statistics corresponding to these M_0 studies be $\{\bar{T}_i\}_{i=1}^{M_0}$. We have

$$\{\bar{T}_i\}_{i=1}^{M_0} \subseteq \{T_i\}_{i=1}^{M} \text{ and } \bar{T}_i \geqslant T_0, i = 1, \ldots, M_0.$$

Notice that $\{\bar{T}_i\}_{i=1}^{M_0}$ are made of two groups of t-statistics. One group includes t-statistics for false discoveries, i.e., discoveries that are insignificant ($\mu_i = 0$) but have a large t-statistic by chance, and the other group includes t-statistics for truly significant discoveries ($\mu_i \neq 0$). Multiple testing aims to develop new threshold values to limit the size of the first group.

In practice, we need to make a choice for T_0. For most applications, it seems that $T_0 = 1.96$ (single test p-value = 5%) is the obvious choice. However, studies with “borderline” t-statistics are difficult to get published. Hence, we will likely have missing observations with statistics around $T_0 = 1.96$. To alleviate this missing data problem, we can choose a higher threshold (e.g., $T_0 = 2.57$, single test p-value = 1%). Correspondingly, we only keep studies in the data that have a t-statistic above this new threshold to construct sample moments for t-statistics. In principle, we can use an even higher threshold value to further reduce the likelihood

[10]See Gerber et al.(2010).

of missing observations. However, this will result in even fewer observations and make the moment estimates less reliable. To strike a balance between bias (i.e., missing observations bias our sample moment estimates) and variability (i.e., fewer observations lead to variable sample moment estimates), we adhere to the somewhat subjective choice of $T_0 = 2.57$.[11]

5.2 Model Estimation

5.2.1 Simulating the Cross-section of Test Statistics

Our estimation first simulates the cross-section of test statistics and then matches key moments of these simulated test statistics to the corresponding moments of the observed test statistics. For a given set of parameters for the structural model, we detail the simulation process as follows.

Step I: Simulate the cross-section of means ($\mu_i, i = 1, \ldots, M$) and persistence parameters ($\nu_i, i = 1, \ldots, M$)

- Generate M values independently from the mixture distribution $p_0 I_{\{\mu=0\}} + (1 - p_0)F(\lambda)$; they correspond to the cross-section of means ($\mu_i, i = 1, \ldots, M$)
- Generate M values independently from the distribution $\Psi(\nu)$; they correspond to the cross-section of persistence parameters ($\nu_i, i = 1, \ldots, M$)

Step II: Simulate a panel of normal shocks $\{W_{ik}, i = 1, \ldots, M, k = 1, \ldots, n\}$

- At each time k, generate a M-dimensional vector from a multivariate normal distribution with mean zero and covariance matrix Φ. The diagonal entries of Φ are $\{(1 - \rho_i^2), i = 1, \ldots, M\}$ and the non-diagonal entries are

[11] In principle, we can estimate the threshold T_0. However, as emphasized by the literature on truncated observations, the estimation for the threshold is unstable. We therefore adopt a pre-selected threshold.

determined by a pre-imposed correlation structure, as parameterized by Φ

- Array these vectors into a $M \times n$ matrix

Step III: Generate the cross-section of t-statistics $\{T_i\}_{i=1}^{M}$

- For each i and with a starting value μ_i^0,[12] construct the time-series $\{U_{ik}, k = 1, \ldots, n\}$ by calculating $U_{ik} = \mu_i(1 - \rho_i) + \rho_i U_{i,k-1} + W_{ik}, k = 1, \ldots, n$ sequentially

- For each i, construct the t-statistic T_i by calculating $T_i = (\sum_{k=1}^{n} U_{ik})/\sqrt{n}$

Step IV: Truncate test statistics at T_0 to obtain the truncated sample $\{\bar{T}\}_{i=1}^{M_0}$

- For each i, keep T_i if it is larger than T_0; the collection of T_i's that are kept is the truncated sample $\{\bar{T}\}_{i=1}^{M_0}$

We repeat the above procedure 10,000 times, each time generating a new sample of t-statistics. We then calculate population moments of test statistics by averaging across these simulated samples.

5.2.2 Estimation

For estimation, we follow the principle of the Generalized Method of Moments (GMM, Hansen, 1982) to match the simulated moments with the sample moments for t-statistics. For the most general model, we collect all model parameters into vector $\Theta = [M, p_0, \lambda', \Psi', \Phi']'$. In Θ, M is the total number of trials, p_0 is the probability of drawing from the null hypothesis, λ parameterizes the distribution F for the non-null hypotheses, Ψ parameterizes the cross-section of persistence parameters

[12] Either a fixed or random seed can be assigned to μ_i^0. We use the unconditional mean μ_i for convenience.

and Φ parameterizes the correlation matrix for the cross-section of contemporaneous shocks.

To understand Θ, we further decompose Θ into two parts:

$$\Theta = [\underbrace{[M, p_0, \lambda']}_{\Theta_1'}, \underbrace{[\Psi', \Phi']}_{\Theta_2'}]'.$$

By this decomposition, Θ_1 includes parameters that traditional methods (e.g., Holm (1979) and Benjamini and Hochberg (1995)) on multiple testing focus on and Θ_2 includes information on correlation or persistence parameters that provide a more detailed description of the panel of shocks.[13] While most studies on multiple testing focus on one or a few components of Θ_1, we are particularly interested in how Θ_2 affects multiple testing adjustments. The inclusion of Θ_2 can change a certain multiple testing adjustment through two channels. First, fixing the estimate of Θ_1, the correlation specification in Θ_2 itself will change the adjustment. In the extreme case when all shocks are perfectly correlated in the cross-section, we do not need any adjustment at all. Second, Θ_2 can affect the adjustment indirectly through Θ_1 as the estimate of Θ_1 depends on the value of Θ_2. Our model thus provides a coherent framework for assessing the impact of Θ_2.

While in principle both Θ_1 and Θ_2 can be estimated within our framework, we choose to estimate Θ_1 only and calibrate Θ_2.[14] Unlike Θ_1, Θ_2 is best thought of as descriptional of the shock structure that generates the cross-section of test statistics. While it could have large impact on multiple testing, we use it primarily as a convenient modeling device to measure the departure from the i.i.d. case for the panel

[13] For instance, to incorporate the information of p_0 into multiple testing, adaptive versions of the Bonferroni and Benjamini-Hochberg methods have been proposed in the literature. See Storey, Taylor and Siegmund (2004), Benjamini, Krieger and Yekutieli (2006), Sarkar (2006, 2008), Gavrilov, Benjamini and Sarkar (2009), Blanchard and Roquain (2008), Finner and Gontscharuk (2009). They try to improve the performances of traditional methods by estimating the number of true null hypotheses.

[14] We discuss several ways to "estimate" Θ_2 at the end of Section 4.

of shocks. More sophisticated modeling of the correlation structure and its impact on multiple testing can be pursued along the lines of the current work but is beyond the scope of the current paper. Additionally, correlations among shocks and hence test statistics are best identified through test statistics that are based on partially overlapped observations.[15] Our framework assumes a balanced panel and should not be used to identify such correlations in estimation.

In sum, we estimate our model and propose multiple testing thresholds for a number of Θ_2 values that can be calibrated. There are several ways to calibrate Θ_2. First, when there is no missing data, ρ can be easily estimated from the average correlation among individual series. As we discussed previously, when there is no missing information, permutation tests bootstrap from individual series to adjust for multiple tests and are robust to non-normality and dependence among test statistics. However, such tests are computationally challenging when the number of tests is large.[16] Our method requires much less computation time and can be valuable in providing a benchmark estimate on what the right adjustment should be.[17] Second, we can estimate the average correlation for a few series and extrapolate for the entire population, providing that we do have information on a few individual series, either through sources that are publicly available or by direct replication of published works. Extrapolation can be dangerous so a range of values for the correlation coefficient should be tried instead of a single point estimate. Lastly, when even a few individual series are difficulty to come by, we can create a table that provides the mapping between the level of correlation and the threshold value for test statistics. Future

[15]See Ferson and Chen (2013) for the same argument on measuring the correlations among mutual fund returns.

[16]For recent works on the permutation approach that try to reduce computational burden, see Lin (2005), Conneely and Boehnke (2007) and Han, Kang and Eskin (2009).

[17]Lin (2005) tries to reduce computation time for permutation tests by simulating innovations from a multivariate normal distribution when the score statistics are known. Our work extends Lin (2005) by also simulating the score statistics.

research can then claim its level of significance under a certain level of correlation.

We estimate Θ_1 by matching moments of the observed and the simulated test statistics. One important moment that we need to include is the total number of published works that have a t-ratio exceeding the pre-specified t-ratio threshold T_0. It helps identify the total number of trials M. Intuitively, fixing all the parameters in Θ_1 except for M, the probability for a test statistic to exceed T_0 is determined. The excepted number of publications is thus proportional to M. Besides this moment, we include the first few sample moments of the observed cross-section of test statistics (e.g., mean, variance, etc.) to estimate other structural parameters. For the simplest model in which Θ_1 is parameterized by three parameters (i.e., λ is a scalar), at least the mean and variance of test statistics are needed to estimate the model. For simplicity, we follow exactly identified GMM to include exactly the same number of moments as the dimension of Θ_1.

5.2.3 Multiple testing adjustment

Finally, we propose multiple testing adjusted thresholds by simulating our structural model at the estimated parameter values. In particular, at a fixed parameter estimate $\hat{\Theta}$ and given a threshold value R for test statistics, we calculate a certain Type I error rate (e.g., FWER or FDR) by simulations. We then search for the optimal threshold value R that achieves a pre-specified significance level.

One important advantage of our structural modeling approach is the precise calculation of Type I/II error rates under various definitions. Unlike a single test, the definition for Type I/II error rates is not straightforward for multiple tests. The statistics literature has proposed several interesting candidates, including FWER and FDR. But there is no consensus as to the preferred choice.[18] Moreover, usual

[18] Alternative definitions akin to FDR include per comparison error rate (Saville, 1990), positive false discovery rate (Storey, 2003) and generalized false discovery rate (Sarkar and Guo, 2009).

adjustments for multiple testing can only achieve a given significance level under certain conditions (e.g., independence among test statistics) and can be very conservative if these conditions are violated. Our simulation framework allows a precise calculation of different forms of error rates. This helps generate multiple testing adjustments that are not overly conservative. In addition, the precise calculation of Type II error rates, however they are defined, allows us to look more into the classic tradeoff between Type I and Type II error rates in a multiple testing context.[19] Our methods offer additional insights on this tradeoff.

5.3 A Simulation Study

5.3.1 Model Simulation

We simulate many t-statistic samples and examine how sample moments change as the parameters for the structural model change. We generate a sample of t-statistics in the following way. We assume that $M = 5,000$ tests have been tried. The test statistics are generated by the structural model at parameter (p_0, λ, ρ), where p_0 is the probability of drawing from the null hypothesis, λ is the single mean parameter that models the exponential distribution for alternative hypotheses and ρ is the correlation coefficient for any pair of normal shocks that constitute the t-statistics. Among these 5,000 trials, only a fraction of them have a t-statistic that is larger than $T_0 = 2.57$. We keep these test statistics and calculate sample statistics based on them. To obtain the distribution of the sample statistics, we repeat the above procedure many times, each time generating a new sample of test statistics. To focus on important parameters, we fix λ at 0.30 and the number of time periods at $N = 240$, i.e., 20 years of monthly data, and vary p_0 and λ to examine their impacts on sample statistics.

[19]Most studies on multiple testing adjustment eschew the issue of Type II errors, mainly because they depend on the high dimensional parameter vector under the alternative hypotheses and are thus difficult to measure.

Table 1 summarizes sample moments for the model simulated at different levels of p_0 and ρ. Fixing ρ, we see that all summary statistics displayed (i.e., mean, median, standard deviation and maximum) of the sample of test statistics are monotonically decreasing in the level of p_0. For example, when $\rho = 0.25$, the 50th percentile of the median of t-statistics changes from 5.76 to 5.38 when p_0 changes from 0.30 to 0.90. Intuitively, this happens because with a higher chance of drawing from the null hypothesis, a higher fraction of the truncated t-statistic sample is made up of false discoveries. This lowers the overall mean and variance of the truncated t-statistic sample as t-statistics under the alternative hypotheses are generally higher and more dispersed than t-statistics under the null hypothesis. The monotone relation between moments of test statistics and p_0 helps identify p_0 when the model is estimated. Lastly, as expected, the number of discoveries M_0 also decreases as p_0 becomes larger. This again can help identify p_0 since the number of discoveries is one of the moments that we need to match in the model estimation.

Fixing p_0, we do not see any noticeable change in sample moments as the level of correlation ρ changes. For instance, fixing p_0 at 0.30, the 50th percentile of the mean of test statistics drops from 7.21 to 7.19 when ρ changes from zero to 0.50. It then increases slightly and reaches 7.20 when ρ further increases to 0.75. None of these changes are important compared to the standard deviation for t-stat means across samples.[20] In contrast, we observe that the variation in M_0 increases dramatically as ρ increases. Fixing p_0 at 0.30, we see that the range between the 10th and 90th percentiles for M_0 changes from 89 (=2123-2034) at $\rho = 0$ to 1023 (=2575-1552) at $\rho = 0.75$. Taken together, correlations among test statistics drive the across-sample variation of the number of discoveries while keeping sample characteristics (i.e., mean,

[20] Based on the range between the 10th and 90th percentiles, the standard deviation for t-stat means across simulated samples is about 0.15 across the four ρ levels when $p_0 = 0.30$.

Table 5.1: **Model Simulation**

Summary statistics for the structural model simulated 5,000 times. The model is parameterized by $\Theta = (M, p_0, \lambda, \rho)'$, where M is the total number of trials and is fixed at 5,000, p_0 is the probability of drawing from the null hypothesis, λ is the mean parameter for the exponential distribution for alternative hypotheses and ρ is the pairwise correlation coefficient between the normal shocks for two different tests. In the table, M_0 reports the number of discoveries (i.e., tests with a t-ratio over 2.57) and "Mean(t-stat)", "Median(t-stat)", "Std(t-stat)" and "Max(t-stat)" report the mean, median, standard deviation and maximum of the sample of test statistics that are above 2.57, respectively.

	$p_0 = 0.30$											
	$\rho = 0$			$\rho = 0.25$			$\rho = 0.50$			$\rho = 0.75$		
Percentiles(%)	10	50	90	10	50	90	10	50	90	10	50	90
M_0	2034	2074	2123	1816	2060	2400	1647	1984	2410	1552	1953	2575
Mean(t-stat)	7.06	7.21	7.32	7.09	7.20	7.32	7.04	7.19	7.30	7.07	7.20	7.37
Median(t-stat)	5.65	5.80	5.90	5.64	5.76	5.89	5.61	5.76	5.88	5.63	5.77	5.93
Std(t-stat)	4.47	4.63	4.84	4.53	4.66	4.83	4.42	4.66	4.86	4.46	4.68	4.83
Max(t-stat)	34.71	39.32	47.17	33.28	39.68	48.91	34.27	39.28	46.40	33.97	39.90	48.09
	$p_0 = 0.60$											
	$\rho = 0$			$\rho = 0.25$			$\rho = 0.50$			$\rho = 0.75$		
Percentiles(%)	10	50	90	10	50	90	10	50	90	10	50	90
M_0	1164	1199	1239	1068	1196	1402	995	1176	1518	942	1179	1468
Mean(t-stat)	6.89	7.10	7.27	6.97	7.18	7.35	6.81	7.14	7.35	6.95	7.17	7.36
Median(t-stat)	5.51	5.68	5.85	5.54	5.74	5.91	5.32	5.75	5.94	5.51	5.78	5.97
Std(t-stat)	4.37	4.59	4.91	4.43	4.66	4.91	4.38	4.61	4.83	4.34	4.62	4.89
Max(t-stat)	31.71	37.02	45.66	32.24	37.58	46.12	31.05	35.34	43.01	32.00	37.16	46.73
	$p_0 = 0.90$											
	$\rho = 0$			$\rho = 0.25$			$\rho = 0.50$			$\rho = 0.75$		
Percentiles(%)	10	50	90	10	50	90	10	50	90	10	50	90
M_0	315	337	360	284	323	415	263	299	446	231	299	403
Mean(t-stat)	6.36	6.65	7.04	6.03	6.73	7.23	5.69	6.99	7.50	5.86	7.08	7.51
Median(t-stat)	4.86	5.17	5.62	3.99	5.38	5.81	3.76	5.55	6.07	4.21	5.64	6.03
Std(t-stat)	4.08	4.58	5.02	3.97	4.49	4.96	4.05	4.54	5.11	4.05	4.53	5.12
Max(t-stat)	24.56	31.23	40.48	24.64	30.20	38.73	25.28	30.51	39.10	24.47	31.15	39.51

median, etc.) largely unchanged. In other words, with high correlations among test statistics, the total number of discoveries made by some literature could be very different — either far above or below its current value — if history repeats. This fact has important implication for the identification of M in the model estimation. Given a high value of ρ, the observed number of discoveries is a noisy indicator of M. Noisy estimates of M in turn imply noisy multiple testing adjustments, as we will see in the next section.

5.3.2 Model Estimation

We now examine the accuracy of the model estimation as well as multiple testing adjustment. Fixing M at 5,000 and λ at 0.30, we simulate one sample of test statistics for a certain combination of p_0 and ρ. We estimate the model based on the simulated sample by pre-specifying a correlation level $\hat{\rho}$, which can be different from the true ρ based on which the data is generated. We denote the estimated parameter vector as $\hat{\Theta}_1 = (\hat{M}, \hat{p_0}, \hat{\lambda})'$. Together with $\hat{\rho}$, we can search for the threshold value $\hat{R}$ that achieves a pre-specified Type I error rate for multiple testing based on the estimated structural model. We use FWER and FDR as the two error rate measures and set the significance level at 10% and 5%, respectively. Lastly, we simulate the true underlying model to calculate the real error rate for the proposed threshold value $\hat{R}$. The resulting real error rates, denoted as $\widehat{\text{FWER}}$ and $\widehat{\text{FDR}}$, should be to close to 10% and 5% respectively if the estimation works well. Therefore, their departure from their target significance levels offers an analytical way to evaluate the performance of our estimation.

Table 2 presents the estimation results. We have several remarks. First, the 50th percentiles of the three parameter estimates are centered around their true values. Notice that this is true even when $\hat{\rho} \neq \rho$, i.e., the specified pairwise correlation differs from the true correlation. This happens because ρ only changes the variation in the number of discoveries across simulated samples, as shown in the previous simulation study. Given a parameter vector $\Theta_1 = (M, p_0, \lambda)'$ and a correlation coefficient ρ, the required sample moments for our estimation (i.e., number of discoveries, mean and variance for test statistics) are almost independent of the level of ρ. This independence results in the unbiasedness in the estimation of Θ_1 even if ρ is misspecified. Second, the two error rates $\widehat{FWER}$ and $\widehat{FDR}$ seem to be estimated unbiasedly only when $\hat{\rho} = \rho$. When $\hat{\rho} > \rho$ ($\hat{\rho} < \rho$), both error rates are overestimated (underesti-

Table 5.2: **Model Estimation for Simulated Samples**

We estimate the structural model parameterized by $\Theta = (M, p_0, \lambda, \rho)'$ 100 times. Each time, we simulate a sample of test statistics based on Θ. Based on this simulated sample and hypothesizing the level of correlation at $\hat{\rho}$, we estimate $\Theta_1 = (M, p_0, \lambda)'$ via GMM. The estimated structural model at $\hat{\Theta} = (\hat{\Theta}_1, \hat{\rho})'$ is then used to generate threshold values that set FWER and FDR at 10% and 5%, respectively. These threshold values are then entered into the true model to generate the estimated FWER ($\widehat{\text{FWER}}$) and FDR ($\widehat{\text{FDR}}$).

								$p_0 = 0.30$								
		$\hat{M}$			$\hat{p}_0$			$\hat{\lambda}$			$\widehat{\text{FWER}}$(%)			$\widehat{\text{FDR}}$(%)		
Percentiles(%)		10	50	90	10	50	90	10	50	90	10	50	90	10	50	90
$\rho = 0$	$\hat{\rho} = 0$	3970	4824	5619	0.20	0.27	0.36	0.21	0.31	0.38	7.34	9.12	13.26	3.78	5.21	7.42
	$\hat{\rho} = 0.25$	3654	4982	5813	0.18	0.27	0.35	0.19	0.28	0.37	8.61	12.82	14.51	4.93	6.69	8.28
	$\hat{\rho} = 0.75$	3691	5012	5942	0.16	0.31	0.41	0.18	0.29	0.42	10.29	14.37	17.83	5.03	7.43	9.91
$\rho = 0.25$	$\hat{\rho} = 0$	2947	4860	6593	0.17	0.22	0.34	0.20	0.33	0.41	5.47	8.19	12.52	2.41	4.29	6.27
	$\hat{\rho} = 0.25$	3089	4851	6531	0.21	0.33	0.39	0.18	0.26	0.35	4.71	10.82	13.96	2.68	4.71	7.03
	$\hat{\rho} = 0.75$	2547	4981	7219	0.22	0.35	0.43	0.16	0.29	0.39	6.47	13.21	17.21	2.51	6.21	8.19
$\rho = 0.75$	$\hat{\rho} = 0$	2868	4789	6342	0.15	0.29	0.45	0.19	0.28	0.37	3.32	7.89	13.21	1.46	3.78	7.02
	$\hat{\rho} = 0.25$	2419	4521	6813	0.24	0.34	0.46	0.18	0.25	0.36	3.98	8.93	14.54	2.31	4.07	7.45
	$\hat{\rho} = 0.75$	2480	4802	6931	0.23	0.35	0.40	0.18	0.29	0.41	4.13	9.89	18.47	2.28	5.43	8.72
								$p_0 = 0.60$								
		$\hat{M}$			$\hat{p}_0$			$\hat{\lambda}$			$\widehat{\text{FWER}}$(%)			$\widehat{\text{FDR}}$(%)		
Percentiles(%)		10	50	90	10	50	90	10	50	90	10	50	90	10	50	90
$\rho = 0$	$\hat{\rho} = 0$	3963	4857	5593	0.49	0.56	0.37	0.22	0.33	0.39	7.74	9.62	13.56	3.38	4.81	7.11
	$\hat{\rho} = 0.25$	3531	5068	5805	0.45	0.56	0.34	0.17	0.25	0.36	8.41	12.72	14.21	4.61	6.44	7.96
	$\hat{\rho} = 0.75$	3678	5117	5873	0.47	0.64	0.43	0.21	0.31	0.41	11.19	15.27	18.73	4.79	7.23	9.37
$\rho = 0.25$	$\hat{\rho} = 0$	3343	4765	5993	0.47	0.54	0.59	0.19	0.32	0.38	5.25	7.99	13.61	2.39	4.17	5.55
	$\hat{\rho} = 0.25$	3172	5050	6497	0.57	0.67	0.73	0.17	0.27	0.37	5.69	10.57	13.03	3.29	4.89	6.71
	$\hat{\rho} = 0.75$	2998	4858	6608	0.59	0.69	0.76	0.13	0.25	0.37	7.36	12.88	16.19	3.99	6.20	7.58
$\rho = 0.75$	$\hat{\rho} = 0$	2853	4794	6401	0.42	0.56	0.73	0.17	0.26	0.37	3.29	7.77	13.04	2.03	3.69	5.87
	$\hat{\rho} = 0.25$	2672	4663	6790	0.51	0.53	0.73	0.20	0.29	0.39	3.85	8.71	14.20	2.29	4.17	6.42
	$\hat{\rho} = 0.75$	3001	4824	6678	0.53	0.63	0.69	0.18	0.26	0.39	5.53	9.87	15.41	3.19	5.44	7.87
								$p_0 = 0.90$								
		$\hat{M}$			$\hat{p}_0$			$\hat{\lambda}$			$\widehat{\text{FWER}}$(%)			$\widehat{\text{FDR}}$(%)		
Percentiles(%)		10	50	90	10	50	90	10	50	90	10	50	90	10	50	90
$\rho = 0$	$\hat{\rho} = 0$	3979	5146	5602	0.76	0.84	0.92	0.25	0.35	0.42	7.96	9.73	13.28	3.56	4.93	6.98
	$\hat{\rho} = 0.25$	3423	4798	5731	0.74	0.86	0.94	0.19	0.24	0.35	8.40	12.58	14.09	4.87	6.05	7.69
	$\hat{\rho} = 0.75$	3578	5231	5871	0.78	0.84	0.93	0.24	0.35	0.43	11.09	14.85	17.67	5.38	7.43	10.03
$\rho = 0.25$	$\hat{\rho} = 0$	2821	4582	6783	0.79	0.87	0.91	0.16	0.30	0.39	6.37	8.14	11.96	2.17	4.29	5.41
	$\hat{\rho} = 0.25$	2628	4967	6892	0.77	0.91	0.95	0.23	0.29	0.36	5.94	10.80	13.43	3.17	4.85	6.69
	$\hat{\rho} = 0.75$	2546	5097	7283	0.75	0.88	0.93	0.16	0.31	0.39	7.74	12.59	15.79	4.03	6.01	7.73
$\rho = 0.75$	$\hat{\rho} = 0$	2268	4795	7097	0.70	0.84	0.93	0.21	0.29	0.36	5.63	7.87	11.79	1.66	3.57	6.05
	$\hat{\rho} = 0.25$	2087	4543	7209	0.78	0.90	0.96	0.25	0.32	0.41	5.61	8.98	13.46	2.48	4.27	6.58
	$\hat{\rho} = 0.75$	2397	5067	7270	0.83	0.92	0.95	0.20	0.25	0.37	6.91	9.37	13.80	2.39	5.30	7.63

mated). Given the unbiased estimate of Θ_1 for the structural model, a conjecture of ρ that is higher than the true ρ (i.e., $\hat{\rho} > \rho$) suggests a threshold value $\hat{R}$ that is lower than what is needed to meet a certain significance level. This explains why both $\widehat{FWER}$ and $\widehat{FDR}$ are above their target levels when $\hat{\rho} > \rho$. In summary, mean estimates imply that while structural parameters are estimated unbiasedly whether the knowledge of ρ is known or not, the accuracy of the error rate estimates crucially depends on a correct specification of ρ.

A closer examination of Table 2 reveals other interesting patterns for our estimation. Irrespective of the level of p_0 and given that ρ is known (i.e., $\hat{\rho} = \rho$), it seems that the standard errors for both parameter and error rate estimates are higher for larger values of ρ. This pattern appears to be more pronounced for M and the two error rates and less so for p_0 and λ. For instance, when $p_0 = 0.60$, the 10%-90% confidence band for the estimate of M is around 2,000 at $\rho = 0$ and increases to 3,000 at $\rho = 0.75$. Similarly, for the estimate of FWER, the confidence band is about 6% at $\rho = 0$ and 10% at $\rho = 0.75$. Again, this happens because the variation in the number of discoveries is large across simulations when ρ is large. Given that M is identified mainly through the number of discoveries, it will vary considerably across simulations. At the same time, error rates crucially depend on the number of trials M. Uncertainty in the estimate of M translates into uncertainty in the estimates of the two error rates. This explains why the standard errors for the error rate estimates are also high when ρ is large.

Lastly, to compare the performance of our model to standard approaches, Table 3 shows the error rate estimates under well-known procedures. In particular, we calculate FWER using Holm (1979)'s procedure and calculate FDR using Benjamini and Hochberg (1995)'s procedure. Also, we separate each error rate calculation into two cases, depending on whether all the tests are observed or not. In the case when not all tests are observed, we assume that only tests with a t-stat over 2.57 are observed. We see from Table 3 that many error rate estimates are very different from their target rates. When all tests are observable, both error rates are close to their targets when $\rho = 0$. This is because both procedures achieve their target rates when tests are independent ($\rho = 0$) and $p_0 = 1$ (i.e., all null hypotheses are true).[21]

[21]See Holm (1979) and Benjamini and Yekutieli (2001) for the proofs.

When ρ gets larger, the two error rates become smaller and further below their target rates. When only significant tests are observed and taken into account in multiple testing adjustment, both methods become too lenient (i.e., the threshold t-statistic is too low), generating error rates that are high above the target rates.

Table 5.3: **Error Rates under Conventional Adjustments**

Type I error rates for the structural model parameterzied by $\Theta = (M, p_0, \lambda, \rho)$, where M and λ are fixed at 5,000 and 0.30, respectively. We simulate a large number of samples of test statistics. For each sample, we apply Holm's adjustment (Holm, 1979) to either the complete sample ("M is known") or the sample truncated at 2.57 ("M is unknown") to generate threshold values. These values are then entered into the underlying true model to determine the true FWER ($\hat{\text{FWER}}$). Analogously, we apply Benjamini and Hochberg's adjustment (Benjamini and Hochberg, 1995) to obtain the true FDR ($\hat{\text{FDR}}$).

	$p_0 = 0.30$											
	M is known						M is unknown					
	$\widehat{\text{FWER}}$(%)			$\widehat{\text{FDR}}$(%)			$\widehat{\text{FWER}}$(%)			$\widehat{\text{FDR}}$(%)		
Percentiles(%)	10	50	90	10	50	90	10	50	90	10	50	90
$\rho = 0$	2.67	3.09	3.87	0.74	1.74	2.24	13.21	15.43	16.21	5.78	9.58	12.29
$\rho = 0.25$	1.49	2.21	2.97	0.77	1.53	1.99	11.54	13.23	15.51	4.65	8.76	12.09
$\rho = 0.75$	0.64	1.73	2.09	0.43	0.96	1.76	8.49	10.29	11.97	2.76	6.53	8.58
	$p_0 = 0.60$											
	M is known						M is unknown					
	$\widehat{\text{FWER}}$(%)			$\widehat{\text{FDR}}$(%)			$\widehat{\text{FWER}}$(%)			$\widehat{\text{FDR}}$(%)		
Percentiles(%)	10	50	90	10	50	90	10	50	90	10	50	90
$\rho = 0$	5.33	5.96	7.34	1.23	3.05	3.97	14.39	16.78	17.59	6.87	10.58	13.39
$\rho = 0.25$	2.20	4.19	5.68	1.59	2.69	3.19	12.69	14.42	16.77	5.93	10.04	13.21
$\rho = 0.75$	1.03	2.48	3.65	0.57	1.05	2.02	9.71	11.79	13.37	3.83	7.46	12.76
	$p_0 = 0.90$											
	M is known						M is unknown					
	$\widehat{\text{FWER}}$(%)			$\widehat{\text{FDR}}$(%)			$\widehat{\text{FWER}}$(%)			$\widehat{\text{FDR}}$(%)		
Percentiles(%)	10	50	90	10	50	90	10	50	90	10	50	90
$\rho = 0$	7.83	8.93	10.47	3.78	4.87	5.31	16.73	19.40	22.77	8.94	13.19	18.41
$\rho = 0.25$	5.67	6.78	8.09	2.87	3.99	4.45	13.47	16.03	19.45	6.95	11.47	16.74
$\rho = 0.75$	2.32	4.39	5.78	1.95	2.58	3.71	10.37	12.96	15.57	4.59	9.94	14.38

Comparing the error rate estimates in Table 2 and 3, our method performs favorably. In particular, when ρ is correctly specified, the error rate estimates based on our structural model are centered around their target rates whereas conventional methods imply estimates that are either too high or too low, depending on whether there is missing data or not. Even when ρ is misspecified, it seems beneficial to have an estimate for the parameters of the structural model. The implied error rates based on the misspecified ρ, albeit biased to some degree, seem to be closer to their target rates than what most conventional methods imply, regardless of whether there is missing data or not.

5.4 Conclusion

In many applications in economics, many candidate variables are used to test a similar hypothesis. For example, there are many papers that study why countries grow at different rates and a large number of "explanatory" variables are proposed. We introduce a new framework that allows for comparisons across many research studies and explicitly controls for the correlation among the proposed variables.

Our approach explicitly models the distributions under the null and alternative hypotheses and decomposes a test statistic into the mean effect and innovations. We simulate the cross-section of innovations and estimate our model using GMM. We show that our estimation works well. We also control publication bias and allow for correlation among tests.

Our research can be enriched in several dimensions. While we assume a correctly specified distributional family for the alternative hypotheses, it would be interesting to see how a misspecified distributional family changes our conclusions. Also, shocks that have fatter tails can be simulated to examine the model's performance. Lastly, from a theoretical point of view, it would be interesting to see how the parameters in our model are exactly identified by the sample moments in the GMM estimation. We leave these issues to future research.

Appendix A

Proofs of Propositions in Chapter 2

A.1 Proof of Proposition 2.

The optimization problem we want to solve is

$$\begin{aligned} \sup_R \quad & E[\frac{R^{1-\gamma}}{1-\gamma}] \\ s.t. \quad & (1)\ E(MR) = 1, \\ & (2)\ R > 0. \end{aligned}$$

For simplicity, I succinctly denote $\gamma(\delta)$ by γ. Moreover, to save space, I only solve the case when $\gamma \in (0,1)$. For $\gamma \in (1,\infty)$ the maximization problem is essentially a minimization problem and a similar proof follows. For $\gamma \in (0,1)$, the maximization problem will be well-defined if all moments of M are assumed to exist. This is because

$$\begin{aligned} E(R^{1-\gamma}) &= E(R^{1-\gamma}M^{1-\gamma}M^{\gamma-1}) \\ &\leqslant [E([(MR)^{1-\gamma}]^{\frac{1}{1-\gamma}})]^{1-\gamma} \cdot [E(M^{\gamma-1})^{\frac{1}{\gamma}}]^{\gamma} \\ &= E(M^{\frac{\gamma-1}{\gamma}})^{\gamma}. \end{aligned}$$

Note that I am using the same trick as in the proof of the new bounds. Also, for $\gamma \in (1,\infty)$ a lower bound for $E(R^{1-\gamma})$ exists so the corresponding minimization problem is also well-defined.

Let the state density function be $f(s)$ and let the Lagrange multipliers associated with $E(MR) = 1$ and $R(s) > 0$ be λ and $\mu(s)$, respectively, then the Lagrange function is

$$\mathcal{L}(R(s), \lambda, \mu(s)) = \frac{1}{1-\gamma} \int R(s)^{1-\gamma} f(s) ds - \lambda(\int M(s)R(s)f(s)ds - 1) - \mu(s)R(s).$$

It is easy to see that the objective function $\frac{1}{1-\gamma} \int R(s)^{1-\gamma} f(s)ds$ is concave in $R(s)$. Additionally, the constraint $\int M(s)R(s)f(s)ds = 1$ is linear in $R(s)$. Under these two conditions, the Kuhn-Tucker first-order conditions are both necessary and sufficient for a maximum of this problem. The first-order condition for the argument $R(s)$ is

$$R(s)^{-\gamma} f(s) - \lambda M(s)f(s) - \mu(s) = 0.$$

Since returns need to have a positive support, the Lagrange multiplier associated with the positivity constraint $\mu(s)$ will be zero in every state. Assuming an everywhere positive $f(s)$, we arrive at the following solution for $R(s)$

$$R(s) = [\lambda(1-\gamma)]^{-\frac{1}{\gamma}} M(s)^{-\frac{1}{\gamma}}. \tag{A.1}$$

To express λ as a moment of the pricing kernel, we can multiply both sides of equation (A.1) by $M(s)f(s)$ and sum across states. This leave us with the following equation for λ

$$[\lambda(1-\gamma)]^{-\frac{1}{\gamma}} E(M^{1-\frac{1}{\gamma}}) = 1. \tag{A.2}$$

Combining equation (A.1) and (A.2), we get the optimal portfolio choice as a function of M only

$$\tilde{R} = M^{-\frac{1}{\gamma}} / E(M^{\frac{\gamma-1}{\gamma}}). \tag{A.3}$$

Note that, by assumption, $M \in Q^{++}$, so $\tilde{R} \in \aleph^{++}$. This validates the earlier step in setting $\mu(s)$ to zero. Finally, by plugging the optimal choice $\tilde{R}$ into the objective function, we have

$$U(M) = \frac{E(\tilde{R}^{1-\gamma})}{1-\gamma} = \frac{[E(M^{\frac{\gamma-1}{\gamma}})]^{\gamma}}{1-\gamma}. \tag{A.4}$$

Equation (A.3) and (A.4) give the optimal solution to this problem.

A.2 Duality definition and proof

For an optimizing investor with a risk-aversion coefficient of γ, her optimization problem is

$$
\begin{aligned}
\sup_R \quad & E[\frac{R^{1-\gamma}}{1-\gamma}] \\
s.t. \quad & (1)\ E(MR) = 1, \\
& (2)\ R > 0.
\end{aligned}
$$

Denote the maximized objective function by $U_{YL}(M)$ and the optimal choice variable by $R_{YL}(M)$, respectively. Note that they are both functionals on M and can be thought of as operators: they operate on any pricing kernel defined on $\aleph^{++}$ and yield a value function and a choice return variable. Symmetrically, a Hansen-Jaganathan type of optimization on the δ-th moment of the pricing kernel can be presented by

$$
\begin{aligned}
\inf_M \quad & \frac{[E(M^\delta)]^{\frac{1}{1-\delta}}}{1-\gamma(\delta)} \\
s.t. \quad & (1)\ E(MR) = 1, \\
& (2)\ M > 0.
\end{aligned}
$$

where $\gamma(\delta) = \frac{1}{1-\delta}$ is what I will term the dual parameter transformation. Similarly, let $U_{HJ}(R)$ and $M_{HJ}(R)$ be the associated functionals (operators). Then a duality between these two optimization problems is satisfied iff the following conditions hold:

$$
\begin{aligned}
R_{YL}(M_{HJ}(R)) &\equiv R \\
M_{HJ}(R_{YL}(M)) &\equiv M.
\end{aligned}
$$

In words, these relationships say the following: 1. The pricing kernel that satisfies the HJ problem with a given return R is the only kernel that can yield an optimal choice of R in my optimization problem; 2. The return that is the optimal choice under my optimization scheme for a given pricking kernel M is the only return that can yield M as the optimal choice in the HJ problem. If two operators satisfy these above duality conditions, then inverse operators can be defined straightforwardly as

$$
\begin{aligned}
R_{YL}^{-1}(R) &\equiv M_{HJ}(R) \\
M_{HJ}^{-1}(M) &\equiv R_{YL}(M).
\end{aligned}
$$

Given the duality definition, it is easy to see that HJ and my optimization are indeed dual problems. To see this, we only need to work out $M_{HJ}(R)$. Similar to Proposition 1, it can be shown that

$$
M_{HJ}(R) = C(R) \cdot R^{\frac{1}{\delta-1}}, \tag{A.5}
$$

where the normalizing constant $C(R)$ is equal to $1/E(R^{\frac{\delta}{\delta-1}})$. By plugging the formulae in equation (A.3) and (A.5) into the duality conditions, it is readily seen that these conditions are satisfied.

A.3 Proof of Proposition 3.

I prove by giving an example. I construct a sequence of pricing kernels that can all price a riskless bond but have either explosive or degenerate δ-th moment in the limit.

Let the state space be $(0, 1)$ and let X be a random variable that is uniformly distributed on $(0, 1)$: $X \sim U(0, 1)$. Let $\{M_n\}_{n=1}^{\infty}$ be a sequence of pricing kernels that are defined by

$$
M_n = \begin{cases} n - \alpha_n & \text{if } X \in (0, \dfrac{1}{n}), \\[2ex] \dfrac{\alpha_n}{n-1} & \text{if } X \in [\dfrac{1}{n}, 1) \end{cases} \tag{A.6}
$$

where $\{\alpha_n\}_{n=1}^{\infty}$ is a sequence that satisfies $\alpha_n < n$ and $\frac{\alpha_n}{n} \to 0$ (For simplicity, α_n can be set at the constant one). Pricing kernels defined in such a way can be understood as describing economies with rare disasters. Rare events happen with a probability $\frac{1}{n}$ and the state price is high in disaster states. Note that a one-period riskless bond has a gross return of one, since $E(M_n) = 1$ for any n. Notice that $E(M_n^{\delta})$ goes to ∞ since

$$
E(M_n^{\delta}) \geqslant (n - \alpha_n)^{\delta} \frac{1}{n} \to \infty
$$

for any $\delta > 1$. However, if a riskless bond is the only security, then return moments are all equal to one. Therefore, no upper bound can be imposed on $E(M^\delta)$. Similarly, for $\delta \in (0, 1)$, we have

$$E(M_n^\delta) = (n - \alpha_n)^\delta \frac{1}{n} + (\frac{\alpha_n}{n-1})^\delta (1 - \frac{1}{n}) \to 0,$$

so no lower bound (except the trivial zero bound) exists for $\delta \in (0, 1)$. Lastly, if $\delta \in (-\infty, 0)$ then no upper bound exists.

Appendix B

Truncated Model Estimation and Bayesian Multiple Testing for Chapter 4

B.1 Multiple Testing When the Number of Tests (M) is Unknown

The empirical difficulty in applying standard p-value adjustments is that we do not observe factors that have been tried, found to be insignificant and then discarded. We attempt to overcome this difficulty using a simulation framework. The idea is first simulate the empirical distribution of p-values for all experiments (published and unpublished) and then adjust p-values based on these simulated samples.

First, we assume the test statistic (t-statistic, for instance) for any experiment follows a certain distribution D (e.g., exponential distribution) and the set of published works is a truncated D distribution. Based on the estimation framework for truncated distributions,[1] we estimate parameters of distribution D and total number of trials M. Next we simulate many sequences of p-values, each corresponding to a plausible set of p-value realizations of all trials. To account for the uncertainty in parameter estimates of D and M, we simulate p-value sequences based on the distribution of estimated D and M. Finally, for each p-value, we calculate the adjusted p-value based on a sequence of simulated p-values. The

[1]See Heckman (1979) and Greene (2008), Chapter 24.

median is taken as the final adjusted p-value.

B.1.1 Using Truncated Exponential Distribution to Model the t-ratio Sample

Truncated distributions have been used to study hidden tests (i.e., publication bias) in medical research.[2] The idea is that studies reporting significant results are more likely to get published. Assuming a threshold significance level or t-statistic, researchers can to some extent infer the results of unpublished works and gain understanding of the overall effect of a drug or treatment. However, in medical research, insignificant results are still viewed as an indispensable part of the overall statistical evidence and are given much more prominence than in the financial economics research. As a result, medical publications tend to report more insignificant results. This makes applying the truncated distribution framework to medical studies difficult as there is no clear-cut threshold value.[3] In this sense, the truncated distributional framework suits our study better — 1.96 is the obvious hurdle that research needs to overcome to get published.

On the other hand, not all tried factors with p-value above 1.96 are reported. In the quantitative asset management industry significant results are not published — they are considered "trade secrets". For the academic literature, factors with "borderline" t-ratios are difficult to get published. Thus, our sample is likely missing a number of factors that have t-ratios just over the bar of 1.96. To make our inference robust, for our baseline result, we assume all tried factors with t-ratios above 2.57 are observed and ignore those with t-ratios in the range of (1.96, 2.57). We experiment with alternative ways to handle t-ratios in this range.

Many distributions can be used to model the t-ratio sample. One restriction that we think any of these distributions should satisfy is the monotonicity of the density curve. Intuitively, it should be easier to find factors with small t-ratios than large ones.[4] We

[2]See Begg and Berlin (1988) and Thornton and Lee (2000).

[3]When the threshold value is unknown, it must be estimated from the likelihood function. However, such estimation usually incurs large estimation errors.

[4]This basic scarcity assumption is also the key ingredient in our model in Section 5.

choose to use the simplest distribution that incorporates this monotonicity condition: the exponential distribution.

Panel A of Figure B.1 presents the histogram of the baseline t-ratio sample and the fitted truncated exponential curve.[5] The fitted density closely tracks the histogram and has a population mean of 2.07.[6] Panel B is a histogram of the original t-ratio sample which, as we discussed before, is likely to under-represent the sample with a t-ratio in the range of (1.96, 2.57). Panel C is the augmented t-ratio sample with the ad hoc assumption that our sample covers only half of all factors with t-ratios between 1.96 and 2.57. The population mean estimate is 2.22 in Panel B and 1.93 in Panel C. As expected, the under-representation of relatively small t-ratios results in a higher mean estimate for the t-ratio population. We think the baseline model is the best among all three models as it not only overcomes the missing data problem for the original sample, but also avoids guessing the fraction of missing observations in the 1.96-2.57 range. We use this model estimates for the follow-up analysis.

Using the baseline model, we calculate other interesting population characteristics that are key to multiple hypothesis testing. Assuming independence, we model observed t-ratios as draws from an exponential distribution with mean parameter $\hat{\lambda}$ and a known cutoff point of 2.57. The proportion of unobserved factors is then estimated as:

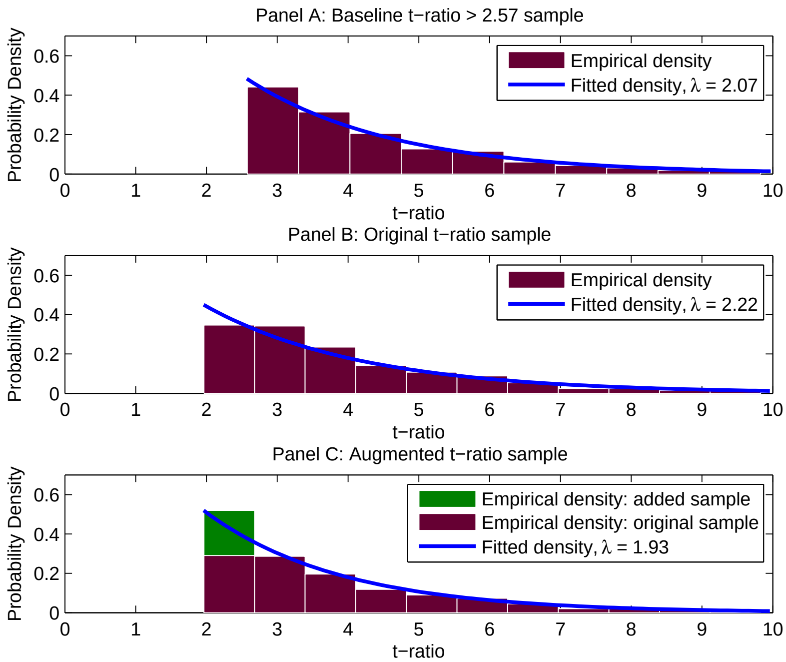
$$P(\text{unobserved}) = \Phi(2.5; \hat{\lambda}) = 1 - \exp(-2.5/\hat{\lambda}) = 71.1\% \tag{B.1}$$

where $\Phi(c; \lambda)$ is the cumulative distribution function evaluated at c for a exponential distribution with mean λ. Our estimates indicate that the mean absolute value of the t-ratio for the underlying factor population is 2.07 and about 71.1% of tried factors are

[5]There are a few very large t-ratios in our sample. We fit the truncated exponential model without dropping any large t-ratios. In contrast to the usual normal density, exponential distribution is better at modeling extreme observations. In addition, extreme values are pivotal statistics for heavy-tailed distributions and are key for model estimation. While extreme observations are included for model estimation, we exclude them in Figure B.1 to better focus on the main part of the t-ratio range.

[6]Our truncated exponential distribution framework allows a simple analytical estimate for the population mean of the exponential distribution. In particular, let c be the truncation point and the t-ratio sample be $\{t_i\}_{i=1}^N$. The mean estimate is given by $\hat{\lambda} = 1/(\bar{t} - c)$, where $\bar{t} = (\sum_{i=1}^N t_i)/N$ is the sample mean.

Figure B.1: **Density Plots for t-ratio**



Empirical density and fitted exponential density curves based on three different samples. Panel A is based on the baseline sample that includes all t-ratios above 2.57. Panel B is based on the original sample with all t-ratios above 1.96. Panel C is based on the augmented sample that adds the sub-sample of observations that fall in between 1.96 and 2.57 to the original t-ratio sample. It doubles the number of observations within the range of 1.96 and 2.57 in the original sample. λ is the single parameter for the exponential curve. It gives the population mean for the unrestricted (i.e., non-truncated) distribution.

discarded. Given that 237 out of the original 315 factors have a t-ratio exceeding 2.57, the total number of factor tests is estimated to be 820 ($= 237/(1 - 71.1\%)$) and the number of factors with a t-ratio between 1.96 and 2.57 is estimated to be 81.[7] Since our t-ratio sample covers only 57 such factors, roughly 30% (=(81-57)/81) of t-ratios between 1.96 and 2.57 are hidden.

[7]Directly applying our estimate framework to the original sample that includes all t-ratios above 1.96, the estimated total number of factor tests would be 719. Alternatively, assuming our sample only covers half of the factors with t-ratios between 1.96 and 2.57, the estimated number of factors is 969.

B.1.2 Simulated Benchmark t-ratios Under Independence

The truncated exponential distribution framework helps us approximate the distribution of t-ratios for all factors, published and unpublished. We can then apply the aforementioned adjustment techniques to this distribution to generate new t-ratio benchmarks. However, there are two sources of sampling and estimation uncertainty that affect our results. First, our t-ratio sample may under-represent all factors with t-statistics exceeding 2.57.[8] Hence, our estimates of total trials are biased (too low), which affects our calculation of the benchmarks. Second, estimation error for the truncated exponential distribution can affect our benchmark t-ratios. Although we can approximate the estimation error through the usual asymptotic distribution theory for MLE, it is unclear how this error affects our benchmark t-ratios. This is because t-ratio adjustment procedures usually depend on the entire t-ratio distribution and so standard transformational techniques (e.g., the delta method) do not apply. Moreover, we are not sure whether our sample is large enough to trust the accuracy of asymptotic approximations.

Given these concerns, we propose a simulation framework that incorporates these uncertainties. We divide it into four steps:

Step I Estimate λ and M based on a new t-ratio sample with size $r \times R$.

Suppose our current t-ratio sample size is R and it only covers a fraction of $1/r$ of all factors. We sample $r \times R$ t-ratios (with replacement) from the original t-ratio sample. Based on this new t-ratio sample, we apply the above truncated exponential distribution framework to the t-ratios and obtain the parameter estimates λ for the exponential distribution. The truncation probability is calculated as $\hat{P} = \Phi(2.5; \hat{\lambda})$. We can then estimate the total number of trials by

$$\hat{M} = \frac{rR}{1 - \hat{P}}$$

[8]This will happen if we miss factors published by the academic literature or we do not have access to the "trade secrets" by industry practitioners.

Step II Calculate the benchmark t-ratio based on a random sample generated from $\hat{\lambda}$ and $\hat{M}$.

Based on the previous step estimate of $\hat{\lambda}$ and $\hat{M}$, we generate a random sample of t-ratios for all tried factors. We then calculate the appropriate benchmark t-ratio based on this generated sample.

Step III Repeat Step II 10,000 times to get the median benchmark t-ratio.

Repeat Step II (based on the same $\hat{\lambda}$ and $\hat{M}$) 10,000 times to generate a collection of benchmark t-ratios. We take the median as the final benchmark t-ratio corresponding to the parameter estimate $(\hat{\lambda}, \hat{M})$.

Step IV Repeat Step I-III 10,000 times to generate a distribution of benchmark t-ratios.

Repeat Step I-III 10,000 times, each time with a newly generated t-ratio sample as in Step I. For each repetition, we obtain a benchmark t-ratio t_i corresponding to the parameter estimates $(\hat{\lambda}_i, \hat{M}_i)$. In the end, we have a collection of benchmark t-ratios $\{t_i\}_{i=1}^{10000}$.

To see how our procedure works, notice that Steps II-III calculate the theoretical benchmark t-ratio for a t-ratio distribution characterized by $(\hat{\lambda}, \hat{M})$. As a result, the outcome is simply one number and there is no uncertainty around it. Uncertainties are incorporated in Steps I and IV. In particular, by sampling repeatedly from the original t-ratio sample and re-estimating λ and M each time, we take into account estimation error of the truncated exponential distribution. Also, under the assumption that neglected significant t-ratios follow the empirical distribution of our t-ratio sample, by varying r, we can assess how this under-representation of our t-ratio sample affects results.

Table B.1 shows estimates of M and benchmark t-ratios. When $r = 1$, the median estimate for the total number of trials is 822,[9] almost the same as our previous estimate

[9] Our previous estimate of 820 is a one-shot estimate based on the truncated sample. The results

of 820 based on the original sample. Unsurprisingly, Bonferroni implied benchmark t-ratio (4.01) is larger than 3.78, which is what we get ignoring unpublished works. Holm implied t-ratio (3.96), while not necessarily increasing in the number of trials, is also higher than before (3.64). BHY implied t-ratio increases from 3.39 to 3.68 at 1% significance and from 2.78 to 3.18 at 5% significance. As r increases, sample size M and benchmark t-ratios for all four types of adjustments increase. When r doubles, the estimate of M also approximately doubles and Bonferroni and Holm implied t-ratios increase by about 0.2, whereas BHY implied t-ratios increase by around 0.03 (under both significance levels).

Table B.1: **Benchmark t-ratios When M is Estimated**

Estimated total number of factors tried (M) and benchmark t-ratio percentiles based on a truncated exponential distribution framework. Our estimation is based on the original t-ratio sample truncated at 2.57. The sampling ratio is the assumed ratio of the true population size of t-ratios exceeding 2.57 over our current sample size. Both Bonferroni and Holm have a significance level of 5%.

Sampling ratio	M	Bonferroni	Holm	BHY(1%)	BHY(5%)
(r)	[10% 90%]	[10% 90%]	[10% 90%]	[10% 90%]	[10% 90%]
1	822	4.01	3.96	3.68	3.18
	[727 937]	[3.98 4.04]	[3.92 4.00]	[3.63 3.74]	[3.12 3.24]
1.5	1229	4.10	4.06	3.70	3.20
	[1125 1370]	[4.08 4.13]	[4.03 4.09]	[3.66 3.75]	[3.16 3.25]
2	1652	4.17	4.13	3.71	3.21
	[1520 1798]	[4.15 4.19]	[4.11 4.16]	[3.67 3.75]	[3.17 3.25]

in Table B.1 are based on repeated estimates based on re-sampled data: we re-sample many times and 822 is the *median* of all these estimates. It is close to the one-shot estimate.

B.2 A Simple Bayesian Framework

The following framework is adopted from Scott and Berger (2006). It highlights the key issues in Bayesian multiple hypothesis testing.[10] More sophisticated generalizations modify the basic model but are unlikely to change the fundamental hierarchical testing structure.[11] We use this framework to explain the pros and cons of performing multiple testing in a Bayesian framework.

The hierarchical model is as follows:

H1. $(X_i|\mu_i, \sigma^2, \gamma_i) \overset{iid}{\sim} N(\gamma_i\mu_i, \sigma^2)$,

H2. $\mu_i|\tau^2 \overset{iid}{\sim} N(0, \tau^2), \gamma_i|p_0 \overset{iid}{\sim} Ber(1 - p_0)$,

H3. $(\tau^2, \sigma^2) \sim \pi_1(\tau^2, \sigma^2), p_0 \sim \pi_2(p_0)$.

We explain each step in detail as well as the notation:

H1. X_i denotes the average return generated from a long-short trading strategy based on a certain factor; μ_i is the unknown mean return; σ^2 is the common variance for returns and γ_i is an indicator function, with $\gamma_i = 0$ indicating a zero factor mean. γ_i is the counterpart of the reject/accept decision in the usual (frequentists') hypothesis testing framework.

H1 therefore says that factor returns are independent conditional on mean $\gamma_i\mu_i$ and common variance σ^2, with $\gamma_i = 0$ indicating that the factor is spurious. The common variance assumption may look restrictive but we can always scale factor returns by changing the dollar investment in the long-short strategy. The crucial assumption

[10] We choose to present the full Bayes approach. An alternative approach — the empirical-Bayes approach — is closely related to the BHY method that controls the *false-discovery rate* (FDR). See Storey (2003) and Efron and Tibshirani (2002) for the empirical-Bayes interpretation of FDR. For details on the empirical-Bayes method, see Efron, Tibshirani, Storey and Tusher (2001), Efron (2004) and Efron (2006). For an in-depth investigation of the differences between the full Bayes and the empirical-Bayes approach, see Scott and Berger (2010).

[11] See Meng and Dempster (1987) and Whittemore (2007) for more works on the Bayesian approach in hypothesis testing.

is conditional independence of average strategy returns. Certain form of conditional independence is unavoidable for Bayesian hierarchical modeling[12] — probably unrealistic for our application. We can easily think of scenarios where average returns of different strategies are correlated, even when population means are known. For example, it is well known that two of the most popular factors, the Fama and French (1992) HML and SMB are correlated.

H2. The first step population parameters μ_i's and γ_i's are assumed to be generated from two other parametric distributions: μ_i's are independently generated from a normal distribution and γ_i's are simply generated from a Bernoulli distribution, i.e., $\gamma_i = 0$ with probability p_0.

The normality assumption for the μ_i's requires the reported X_i's to randomly represent either long/short or short/long strategy returns. If researchers have a tendency to report positive abnormal returns, we need to randomly assign to these returns plus/minus signs. The normality assumptions in both H1 and H2 are important as they are necessary to guarantee the properness of the posterior distributions.

H3. Finally, the two variance variables τ^2 and σ^2 follow a joint prior distribution π_1 and the probability p_0 follows a prior distribution π_2.

Objective or "neutral" priors for π_1 and π_2 can be specified as:

$$\begin{aligned} \pi_1(\tau^2, \sigma^2) &\propto (\tau^2 + \sigma^2)^{-2} \\ \pi_2(p_0) &= \text{Uniform}(0, 1) \end{aligned}$$

Under this framework, the joint conditional likelihood function for X_i's is simply a product of individual normal likelihood functions and the posterior probability that $\gamma_i = 1$ (discovery) can be calculated by applying Bayes' law. When the number of trials is large, to

[12] Conditional independence is crucial for the Bayesian framework and the construction of posterior likelihoods. Although it can be extended to incorporate special dependence structures, there is no consensus on how to systematically handle dependence. See Brown et al. (2012) for a discussion of independence in Bayesian multiple testing. They also propose a spatial dependence structure into a Bayesian testing framework.

calculate the posterior probability we need efficient methods such as importance sampling, which involves high dimensional integrals.

One benefit of a Bayesian framework for multiple testing is that the multiplicity penalty term is already embedded. In the frequentists' framework, this is done by introducing FWER or FDR. In a Bayesian framework, the so-called "Ockham's razor effect"[13] automatically adjusts the posterior probabilities when more factors are simultaneously tested.[14] Simulation studies in Scott and Berger (2006) show how the discovery probabilities for a few initial signals increase when more noise are added to the original sample.

However, there are several shortcomings for the Bayesian approach. Some of them are specific to the context of our application and the others are generic to the Bayesian multiple testing framework.

At least two issues arise when applying the Bayesian approach to our factor selection problem. First, we do not observe all tried factors. While we back out the distribution of hidden factors parametrically under the frequentist framework, it is not clear how the missing data and the multiple testing problems can be simultaneously solved under the Bayesian framework. Second, the hierarchical testing framework may be overly restrictive. Both independence as well as normality assumptions can have a large impact on the posterior distributions. Although normality can be somewhat relaxed by using alternative distributions, the scope of alternative distributions is limited as there are only a few distributions that can guarantee the properness of the posterior distributions. Independence, as we previously discussed, is likely to be violated in our context. In contrast, the three adjustment procedures under the frequentists' framework are able to handle complex data structures since they rely on only fundamental probability inequalities to restrict their objective function — the Type I error rate.

There are a few general concerns about the Bayesian multiple testing framework. First,

[13]See Jefferys and Berger (1992).

[14]Intuitively, more complex models are penalized because extra parameters involve additional sources of uncertainty. Simplicity is rewarded in a Bayesian framework as simple models produce sharp predictions. See the discussions in Scott (2009).

it is not clear what to do after obtaining the posterior probabilities for individual hypotheses. Presumably, we should find a cutoff probability P and reject all hypotheses that have a posterior discovery probability larger than P. But then we come back to the initial problem of finding an appropriate cutoff p-value, which is not at all a clear task. Scott and Berger (2006) suggest a decision-theoretic approach that chooses the cutoff P by minimizing a loss-function. The parameters of the loss-function, however, are again subjective. Second, the Bayesian posterior distributions are computationally challenging. We document three hundred factors but there are potentially many more if missing factors are taken into account. When M gets large, importance sampling is a necessity. However, results of importance sampling rely on simulations and subjective choices of the centers of the probability distributions for random variables. Consequently, two researchers trying to calculate the same quantity might get very different results. Moreover, in multiple testing, the curse of dimensionality generates additional risks for Bayesian statistical inference.[15] These technical issues create additional hurdles for the application of the Bayesian approach.

[15] See Liang and Kelemen (2008) for a discussion on the computational issues in Bayesian multiple testing.

B.3 Method Controlling the FDP

We apply the methods developed in Lehmann and Romano (2005) to control the realized FDP. The objective is $P(FDP > \gamma) \leqslant \alpha$, where γ is the threshold FDP value and α is the significance level. Fixing γ and α, we order the individual p-values from the smallest to the largest (i.e., $p_{(1)} \leqslant p_{(2)} \leqslant \cdots \leqslant p_{(M)}$) and let the corresponding hypotheses be $H_{(1)}, H_{(2)}, \cdots, H_{(M)}$. We then reject the i-th hypothesis if $p_{(i)} \leqslant \alpha_i / C_{\lfloor \gamma M \rfloor + 1}$, where

$$\alpha_i = \frac{(\lfloor \gamma i \rfloor + 1)\alpha}{M + \lfloor \gamma i \rfloor + 1 - i},$$

$$C_k = \sum_{j=1}^{k} \frac{1}{j}.$$

Here, for a real number x, $\lfloor x \rfloor$ denotes the greatest integer that is no greater than x. Similar to $c(M)$ in BHY's adjustment, $C_{\lfloor \gamma M \rfloor + 1}$ allow one to control the FDP under arbitrary dependence structure of the p-values.

Table B.2 shows the benchmark t-ratios based on our sample of 315 factors for different levels of FDP thresholds and significance . The benchmark t-ratios are higher when the FDP thresholds are tougher (i.e., γ is lower) or when the significance levels are lower (i.e., α is lower). For typical values of γ and α, the benchmark t-ratios are significantly lower than conventional values, consistent with previous results based on the FWER or FDR methods. For instance, when $\gamma = 0.10$ and $\alpha = 0.05$, the benchmark t-ratio is 2.70 (p-value $= 0.69\%$), much higher than the conventional cutoff of 1.96.

Table B.2: **Benchmark t-ratios for Lehmann and Romano (2005)**

Estimated benchmark t-ratios based on Lehmann and Romano (2005). The objective is $P(FDP > \gamma) \leqslant \alpha$.

	$\gamma = 0.05$	$\gamma = 0.10$	$\gamma = 0.20$
$\alpha = 0.01$	3.70	3.48	3.25
$\alpha = 0.05$	3.04	2.70	2.38
$\alpha = 0.10$	2.38	2.17	2.16

Bibliography

Aad, G., Abajyan, T., Abbott, B., Abdallah, J., Abdel Khalek, S., Abdelalim, A., Abdinov, O., Aben, R., Abi, B., Abolins, M., et al. (2012), "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Physics Letters B*, 716, 1–29.

Abarbanell, J. S. and Bushee, B. J. (1998), "Abnormal returns to a fundamental analysis strategy," *Accounting Review*, pp. 19–45.

Acharya, V. V. and Pedersen, L. H. (2005), "Asset pricing with liquidity risk," *Journal of Financial Economics*, 77, 375–410.

Ackert, L. F. and Athanassakos, G. (1997), "Prior uncertainty, analyst bias, and subsequent abnormal returns," *Journal of Financial Research*, 20, 263–273.

Adler, M. and Solnik, B. (1974), "The international pricing of risk: an empirical investigation of the world capital market structure," *The Journal of Finance*, 29, 365–378.

Adrian, T. and Rosenberg, J. (2008), "Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk," *The Journal of Finance*, 63, 2997–3030.

Adrian, T., Etula, E., and Muir, T. (2012), "Financial intermediaries and the cross-section of asset returns," *Staff Reports*, 464.

Ahn, S. C., Horenstein, A. R., and Wang, N. (2012), "Determining rank of the beta matrix of a linear asset pricing model," Tech. rep., Working Paper, Arizona State University and Sogang University.

Ai, H. (2010), "Information Quality and Long-Run Risk: Asset Pricing Implications," *The Journal of Finance*, 65, 1333–1367.

Aıt-Sahalia, Y., Wang, Y., and Yared, F. (2001), "Do option markets correctly price the probabilities of movement of the underlying asset?" *Journal of Econometrics*, 102, 67–110.

Akbas, F., Armstrong, W. J., Sorescu, S. M., and Subrahmanyam, A. (2013), "Time varying market efficiency in the cross-section of expected stock returns," in *AFA 2013 San Diego Meetings Paper*.

Ali, A., Hwang, L.-S., and Trombley, M. A. (2003), "Arbitrage risk and the book-to-market anomaly," *Journal of Financial Economics*, 69, 355–373.

Almeida, C. and Garcia, R. (2012), "Assessing misspecified asset pricing models with empirical likelihood estimators," *Journal of Econometrics*, 170, 519–537.

Almeida, C. and Garcia, R. (2013), "Robust Assessment of Hedge Fund Performance through Nonparametric Risk Adjustment," *Available at SSRN*.

Almeida, H. and Campello, M. (2007), "Financial constraints, asset tangibility, and corporate investment," *Review of Financial Studies*, 20, 1429–1460.

Alvarez, F. and Jermann, U. J. (2005), "Using asset prices to measure the persistence of the marginal utility of wealth," *Econometrica*, 73, 1977–2016.

Amaya, D., Christoffersen, P., Jacobs, K., and Vasquez, A. (2011), "Do realized skewness and kurtosis predict the cross-section of equity returns," Tech. rep., School of Economics and Management, University of Aarhus.

Amihud, Y. (2002), "Illiquidity and stock returns: cross-section and time-series effects," *Journal of financial markets*, 5, 31–56.

Amihud, Y. and Mendelson, H. (1986), "Asset pricing and the bid-ask spread," *Journal of financial Economics*, 17, 223–249.

Amihud, Y. and Mendelson, H. (1989), "The Effects of Beta, Bid-Ask Spread, Residual Risk, and Size on Stock Returns," *The Journal of Finance*, 44, 479–486.

An, B.-J., Ang, A., Bali, T. G., and Cakici, N. (2014), "The joint cross section of stocks and options," *The Journal of Finance*.

An, J., Bhojraj, S., and Ng, D. T. (2010), "Warranted multiples and future returns," *Journal of Accounting, Auditing & Finance*, 25, 143–169.

Anderson, C. W. and GARCIA-FEIJÓO, L. (2006), "Empirical evidence on capital investment, growth options, and security returns," *The Journal of Finance*, 61, 171–194.

Anderson, E. W., Ghysels, E., and Juergens, J. L. (2005), "Do heterogeneous beliefs matter for asset pricing?" *Review of Financial Studies*, 18, 875–924.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006a), "The cross-section of volatility and expected returns," *The Journal of Finance*, 61, 259–299.

Ang, A., Chen, J., and Xing, Y. (2006b), "Downside risk," *Review of Financial Studies*, 19, 1191–1239.

Arbel, A., Carvell, S., and Strebel, P. (1983), "Giraffes, institutions and neglected firms," *Financial Analysts Journal*, pp. 57–63.

Armstrong, C., Banerjee, S., and Corona, C. (2009), "Information quality and the cross-section of expected returns," in *AFA 2010 Meeting Paper (Working Paper)*.

Asness, C. S., Porter, R. B., and Stevens, R. L. (2000), "Predicting stock returns using industry-relative firm characteristics," *Available at SSRN 213872*.

Asquith, P., Pathak, P. A., and Ritter, J. R. (2005), "Short interest, institutional ownership, and stock returns," *Journal of Financial Economics*, 78, 243–276.

Avramov, D., Chordia, T., Jostova, G., and Philipov, A. (2007), "Momentum and credit rating," *The Journal of Finance*, 62, 2503–2520.

Avramov, D., Chordia, T., Jostova, G., and Philipov, A. (2009), "Dispersion in analysts earnings forecasts and credit rating," *Journal of Financial Economics*, 91, 83–101.

Backus, D., Chernov, M., and Martin, I. (2011), "Disasters implied by equity index options," *The journal of finance*, 66, 1969–2012.

Backus, D., Chernov, M., and Zin, S. (2014), "Sources of entropy in representative agent models," *The Journal of Finance*, 69, 51–99.

Baik, B. and Ahn, T. (2007), "Changes in order backlog and future returns," .

Bajgrowicz, P. and Scaillet, O. (2012), "Technical trading revisited: False discoveries, persistence tests, and transaction costs," *Journal of Financial Economics*, 106, 473–491.

Bajgrowicz, P., Scaillet, O., and Treccani, A. (2011), "Jumps in high-frequency data: Spurious detections, dynamics, and news," *Swiss Finance Institute Research Paper*.

Baker, M. and Wurgler, J. (2006), "Investor sentiment and the cross-section of stock returns," *The Journal of Finance*, 61, 1645–1680.

Bakshi, G. and Kapadia, N. (2003), "Delta-hedged gains and the negative market volatility risk premium," *Review of Financial Studies*, 16, 527–566.

Bakshi, G., Cao, C., and Chen, Z. (1997), "Empirical performance of alternative option pricing models," *The Journal of Finance*, 52, 2003–2049.

Balachandran, S. and Mohanram, P. (2012), "Using residual income to refine the relationship between earnings growth and stock returns," *Review of Accounting Studies*, 17, 134–165.

Balduzzi, P. and Robotti, C. (2008), "Mimicking portfolios, economic risk premia, and tests of multi-beta models," *Journal of Business & Economic Statistics*, 26, 354–368.

Bali, T. and Zhou, H. (2012), "Risk, uncertainty, and expected returns," *Available at SSRN 2020604*.

Bali, T. G. and Hovakimian, A. (2009), "Volatility spreads and expected stock returns," *Management Science*, 55, 1797–1812.

Bali, T. G., Cakici, N., and Whitelaw, R. F. (2011), "Maxing out: Stocks as lotteries and the cross-section of expected returns," *Journal of Financial Economics*, 99, 427–446.

Baltussen, G., Van Bekkum, S., and Van Der Grient, B. (2013), "Unknown Unknowns: Vol-of-Vol and the Cross-Section of Stock Returns," in *AFA 2013 San Diego Meetings Paper*.

Balvers, R. J. and Huang, D. (2007), "Productivity-based asset pricing: Theory and evidence," *Journal of Financial Economics*, 86, 405–445.

Bandyopadhyay, S. P., Huang, A. G., and Wirjanto, T. S. (2010), "The accrual volatility anomaly," Tech. rep., working paper, University of Waterloo.

Bansal, R. and Lehmann, B. N. (1997), "Growth-optimal portfolio restrictions on asset pricing models," *Macroeconomic Dynamics*, 1, 333–354.

Bansal, R. and Shaliastovich, I. (2011), "Learning and asset-price jumps," *Review of Financial Studies*, p. hhr023.

Bansal, R. and Viswanathan, S. (1993), "No arbitrage and arbitrage pricing: A new approach," *The Journal of Finance*, 48, 1231–1262.

Bansal, R. and Yaron, A. (2004), "Risks for the long run: A potential resolution of asset pricing puzzles," *The Journal of Finance*, 59, 1481–1509.

Bansal, R., Dittmar, R. F., and Lundblad, C. T. (2005), "Consumption, dividends, and the cross section of equity returns," *The Journal of Finance*, 60, 1639–1672.

Bansal, R., Kiku, D., and Yaron, A. (2009), "An empirical evaluation of the long-run risks model for asset prices," Tech. rep., National Bureau of Economic Research.

Banz, R. W. (1981), "The relationship between return and market value of common stocks," *Journal of financial economics*, 9, 3–18.

Barber, B., Lehavy, R., McNichols, M., and Trueman, B. (2001), "Can investors profit from the prophets? Security analyst recommendations and stock returns," *The Journal of Finance*, 56, 531–563.

Barber, B. M., Odean, T., and Zhu, N. (2009), "Do retail trades move markets?" *Review of Financial Studies*, 22, 151–186.

Barras, L., Scaillet, O., and Wermers, R. (2010), "False discoveries in mutual fund performance: Measuring luck in estimated alphas," *The Journal of Finance*, 65, 179–216.

Barro, R. J. and Ursúa, J. F. (2009), "Stock-market crashes and depressions," Tech. rep., National Bureau of Economic Research.

Barro, R. J. and Ursúa, J. F. (2011), "Rare macroeconomic disasters," Tech. rep., National Bureau of Economic Research.

Basu, S. (1977), "Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis," *The Journal of Finance*, 32, 663–682.

Basu, S. (1983), "The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence," *Journal of financial economics*, 12, 129–156.

Bauman, W. S. and Dowen, R. (1988), "Growth projections and common stock returns," *Financial Analysts Journal*, 44, 79–80.

Begg, C. B. and Berlin, J. A. (1988), "Publication bias: a problem in interpreting medical data," *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pp. 419–463.

Bekaert, G. and Liu, J. (2004), "Conditioning information and variance bounds on pricing kernels," *Review of Financial Studies*, 17, 339–378.

Belo, F., Lin, X., and Bazdresch, S. (2014), "Labor hiring, investment, and stock return predictability in the cross section," *Journal of Political Economy*, 122, 129–177.

Beneish, M. D. (1997), "Detecting GAAP violation: Implications for assessing earnings management among firms with extreme financial performance," *Journal of Accounting and Public Policy*, 16, 271–309.

Beneish, M. D., Lee, M., and Nichols, D. C. (2012), "Fraud detection and expected returns," *Available at SSRN.*

Benjamini, Y. and Hochberg, Y. (1995), "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 289–300.

Benjamini, Y. and Liu, W. (1999), "A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence," *Journal of Statistical Planning and Inference*, 82, 163–170.

Benjamini, Y. and Yekutieli, D. (2001), "The control of the false discovery rate in multiple testing under dependency," *Annals of statistics*, pp. 1165–1188.

Berkman, H., Jacobsen, B., and Lee, J. B. (2011), "Time-varying rare disaster risk and stock returns," *Journal of Financial Economics*, 101, 313–332.

Bhandari, L. C. (1988), "Debt/equity ratio and expected common stock returns: Empirical evidence," *The Journal of Finance*, 43, 507–528.

Black, F. (1972), "Capital market equilibrium with restricted borrowing," *Journal of business*, pp. 444–455.

Boguth, O. and KUEHN, L.-A. (2013), "Consumption volatility risk," *The Journal of Finance*, 68, 2589–2615.

Bondarenko, O. (2003), "Why are put options so expensive," *University of Illinois at.*

Bondt, W. F. and Thaler, R. (1985), "Does the stock market overreact?" *The Journal of finance*, 40, 793–805.

Boons, M., De Roon, F., and Szymanowska, M. (2012), "The stock market price of commodity risk," in *AFA 2012 Chicago Meetings Paper*.

Bossaerts, P. and Dammon, R. M. (1994), "Tax-Induced Intertemporal Restrictions on Security Returns," *The Journal Of Finance*, 49, 1347–1371.

Botosan, C. A. (1997), "Disclosure level and the cost of equity capital," *Accounting review*, pp. 323–349.

Boudoukh, J., Michaely, R., Richardson, M., and Roberts, M. R. (2007), "On the importance of measuring payout yield: Implications for empirical asset pricing," *The Journal of Finance*, 62, 877–915.

Boyer, B., Mitton, T., and Vorkink, K. (2009), "Expected idiosyncratic skewness," *Review of Financial Studies*, p. hhp041.

Bradshaw, M. T., Richardson, S. A., and Sloan, R. G. (2006), "The relation between corporate financing activities, analysts forecasts and stock returns," *Journal of Accounting and Economics*, 42, 53–85.

Brammer, S., Brooks, C., and Pavelin, S. (2006), "Corporate social performance and stock returns: UK evidence from disaggregate measures," *Financial Management*, 35, 97–116.

Brandt, M. W. (1999), "Estimating portfolio and consumption choice: A conditional Euler equations approach," *The Journal of Finance*, 54, 1609–1645.

Brandt, M. W. and Santa-Clara, P. (2006), "Dynamic portfolio selection by augmenting the asset space," *The Journal of Finance*, 61, 2187–2217.

Brandt, M. W., Kishore, R., Santa-Clara, P., and Venkatachalam, M. (2008), "Earnings announcements are full of surprises," *SSRN eLibrary*.

Breeden, D. T. (1979), "An intertemporal asset pricing model with stochastic consumption and investment opportunities," *Journal of financial Economics*, 7, 265–296.

Breeden, D. T., Gibbons, M. R., and Litzenberger, R. H. (1989), "Empirical tests of the consumption-oriented CAPM," *The Journal of Finance*, 44, 231–262.

Brennan, M., Huh, S.-W., and Subrahmanyam, A. (2013), "The pricing of good and bad private information in the cross-section of expected stock returns," Tech. rep., Working Paper, University of California at Los Angeles.

Brennan, M. J. and Li, F. (2008), "Agency and asset pricing," *Available at SSRN 1104546*.

Brennan, M. J. and Subrahmanyam, A. (1996), "Market microstructure and asset pricing: On the compensation for illiquidity in stock returns," *Journal of financial economics*, 41, 441–464.

Brennan, M. J., Chordia, T., and Subrahmanyam, A. (1998), "Alternative factor specifications, security characteristics, and the cross-section of expected stock returns," *Journal of Financial Economics*, 49, 345–373.

Brennan, M. J., Wang, A. W., and Xia, Y. (2004), "Estimation and test of a simple model of intertemporal capital asset pricing," *The Journal of Finance*, 59, 1743–1776.

Brennan, M. J., Chordia, T., Subrahmanyam, A., and Tong, Q. (2012), "Sell-order liquidity and the cross-section of expected stock returns," *Journal of Financial Economics*, 105, 523–541.

Broadie, M., Chernov, M., and Johannes, M. (2009), "Understanding index option returns," *Review of Financial Studies*, 22, 4493–4529.

Brown, D. A., Lazar, N. A., Datta, G. S., Jang, W., and McDowell, J. E. (2014), "Incorporating spatial dependence into Bayesian multiple testing of statistical parametric maps in functional neuroimaging," *NeuroImage*, 84, 97–112.

Brown, D. P. and Rowe, B. (2007), "The productivity premium in equity returns," *Available at SSRN 993467*.

Buraschi, A. and Jackwerth, J. (2001), "The price of a smile: Hedging and spanning in option markets," *Review of Financial Studies*, 14, 495–527.

Burlacu, R., Fontaine, P., Jimenez-Garces, S., and Seasholes, M. S. (2012), "Risk and the cross section of stock returns," *Journal of Financial Economics*, 105, 511–522.

Callen, J. L. and Lyle, M. R. (2014), "The term structure of implied costs of equity capital," *Rotman School of Management Working Paper*.

Callen, J. L., Khan, M., and Lu, H. (2013), "Accounting Quality, Stock Price Delay, and Future Stock Returns*," *Contemporary Accounting Research*, 30, 269–295.

Campbell, J. Y. (1993), "Understanding risk and return," Tech. rep., National Bureau of Economic Research.

Campbell, J. Y. and Cochrane, J. H. (1995), "By force of habit: A consumption-based explanation of aggregate stock market behavior," Tech. rep., National Bureau of Economic Research.

Campbell, J. Y. and Vuolteenaho, T. (2003), "Bad beta, good beta," Tech. rep., National Bureau of Economic Research.

Campbell, J. Y., Hilscher, J., and Szilagyi, J. (2008), "In search of distress risk," *The Journal of Finance*, 63, 2899–2939.

Campbell, J. Y., Giglio, S., Polk, C., and Turley, R. (2012), "An intertemporal capm with stochastic volatility," Tech. rep., National Bureau of Economic Research.

Cao, C., Chen, Y., Liang, B., and Lo, A. W. (2013), "Can hedge funds time market liquidity?" *Journal of Financial Economics*, 109, 493–516.

Cao, X. and Xu, Y. (2010), "Long-run idiosyncratic volatilities and cross-sectional stock returns," *Available at SSRN 1569945*.

Carhart, M. M. (1997), "On persistence in mutual fund performance," *The Journal of finance*, 52, 57–82.

Cen, L., Wei, J., and Zhang, J. (2006), "Forecasted earnings per share and the cross section of expected stock returns," Tech. rep., Working Paper, Hong Kong University of Science & Technology.

Chan, K., Chen, N.-f., and Hsieh, D. A. (1985), "An exploratory investigation of the firm size effect," *Journal of Financial Economics*, 14, 451–471.

Chan, K., Foresi, S., and Lang, L. H. (1996), "Does money explain asset returns? Theory and empirical analysis," *The Journal of Finance*, 51, 345–361.

Chan, Y. L. and Kogan, L. (2001), "Catching up with the Joneses: Heterogeneous preferences and the dynamics of asset prices," Tech. rep., National Bureau of Economic Research.

Chandrashekar, S. and Rao, R. K. (2009), "The productivity of corporate cash holdings and the cross-section of expected stock returns," *McCombs Research Paper Series No. FIN-03-09*.

Chang, B. Y., Christoffersen, P., and Jacobs, K. (2013), "Market skewness risk and the cross section of stock returns," *Journal of Financial Economics*, 107, 46–68.

Chapman, D. A. (1997), "Approximating the asset pricing kernel," *The Journal of Finance*, 52, 1383–1410.

Chatrchyan, S., Khachatryan, V., Sirunyan, A. M., Tumasyan, A., Adam, W., Aguilo, E., Bergauer, T., Dragicevic, M., Erö, J., Fabjan, C., et al. (2012), "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Physics Letters B*, 716, 30–61.

Chemmanur, T. J. and Yan, A. (2009), "Advertising, attention, and stock returns," *Attention, and Stock Returns (February 10, 2009)*.

Chen, H. J., Kacperczyk, M., and Ortiz-Molina, H. (2011), "Labor unions, operating flexibility, and the cost of equity," *Journal of Financial and Quantitative Analysis*, 46, 25–58.

Chen, J., Hong, H., and Stein, J. C. (2002), "Breadth of ownership and stock returns," *Journal of financial Economics*, 66, 171–205.

Chen, L., Novy-Marx, R., and Zhang, L. (2010), "An alternative three-factor model," *Unpublished paper, Washington University in St. Louis, University of Chicago, and University of Michigan*.

Chen, N.-F., Roll, R., and Ross, S. A. (1986), "Economic forces and the stock market," *Journal of business*, pp. 383–403.

Chen, Z. and Petkova, R. (2012), "Does idiosyncratic volatility proxy for risk exposure?" *Review of Financial Studies*, 25, 2745–2787.

Chen, Z. and Strebulaev, I. (2013), "Contingent-claim-based expected stock returns," *Available at SSRN 2018320*.

Chopra, N., Lakonishok, J., and RITTER, J. (1993), "Measuring Abnormal Performance," *Advances in behavioral finance*, 1, 265.

Chordia, T. and Shivakumar, L. (2006), "Earnings and price momentum," *Journal of financial economics*, 80, 627–656.

Chordia, T., Subrahmanyam, A., and Anshuman, V. R. (2001), "Trading activity and expected stock returns," *Journal of Financial Economics*, 59, 3–32.

Chordia, T., Huh, S.-W., and Subrahmanyam, A. (2009), "Theory-based illiquidity and asset pricing," *Review of Financial Studies*, 22, 3629–3668.

Chordia, T., Subrahmanyam, A., and Tong, Q. (2013), "Trends in capital market anomalies," *Unpublished working paper, Emory University, UCLA, Singapore Management University*.

Chung, Y. P., Johnson, H., and Schill, M. J. (2006), "Asset Pricing When Returns Are Nonnormal: Fama-French Factors versus Higher-Order Systematic Comoments*," *The Journal of Business*, 79, 923–940.

Cochrane, J. H. (1991), "Production-based asset pricing and the link between stock returns and economic fluctuations," *The Journal of Finance*, 46, 209–237.

Cochrane, J. H. (1996), "A cross-sectional test of a production-based asset pricing model," Tech. rep., National Bureau of Economic Research.

Cochrane, J. H. (2011), "Presidential address: Discount rates," *The Journal of Finance*, 66, 1047–1108.

Cohen, L. and Frazzini, A. (2008), "Economic links and predictable returns," *The Journal of Finance*, 63, 1977–2011.

Cohen, L. and Lou, D. (2012), "Complicated firms," *Journal of financial economics*, 104, 383–400.

Cohen, L., Malloy, C., and Pomorski, L. (2012), "Decoding inside information," *The Journal of Finance*, 67, 1009–1043.

Cohen, L., Diether, K., and Malloy, C. (2013), "Misvaluing innovation," *Review of Financial Studies*, p. hhs183.

Cohen, R., Polk, C., and Silli, B. (2010), "Best ideas," *Unpublished working paper. Harvard Business School*.

Colacito, R., Ghysels, E., and Meng, J. (2013), "Skewness in expected macro fundamentals and the predictability of equity returns: Evidence and theory," *Unpublished working paper. University of North Carolina*.

Conrad, J., Cooper, M., and Kaul, G. (2003), "Value versus glamour," *The Journal of Finance*, 58, 1969–1996.

Conrad, J., Dittmar, R. F., and Ghysels, E. (2013), "Ex ante skewness and expected stock returns," *The Journal of Finance*, 68, 85–124.

Constantinides, G. M. (1982), "Intertemporal asset pricing with heterogeneous consumers and without demand aggregation," *Journal of Business*, pp. 253–267.

Constantinides, G. M. (1986), "Capital market equilibrium with transaction costs," *The Journal of Political Economy*, pp. 842–862.

Cooper, M. J., Gulen, H., and Schill, M. J. (2008), "Asset growth and the cross-section of stock returns," *The Journal of Finance*, 63, 1609–1651.

Cooper, M. J., Gulen, H., and Ovtchinnikov, A. V. (2010), "Corporate political contributions and stock returns," *The Journal of Finance*, 65, 687–724.

Coval, J. D. and Shumway, T. (2001), "Expected option returns," *The journal of Finance*, 56, 983–1009.

Cox, D. (1982), "Statistical significance tests." *British journal of clinical pharmacology*, 14, 325–331.

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985), "An intertemporal general equilibrium model of asset prices," *Econometrica: Journal of the Econometric Society*, pp. 363–384.

Cremers, K. and Nair, V. B. (2005), "Governance mechanisms and equity prices," *The Journal of Finance*, 60, 2859–2894.

Cremers, K. M., Nair, V. B., and John, K. (2009), "Takeovers and the cross-section of returns," *Review of Financial Studies*, 22, 1409–1445.

Cremers, M., Halling, M., and Weinbaum, D. (2010), "In search of aggregate jump and volatility risk in the cross-section of stock returns," *SSRN eLibrary*.

Da, Z. (2009), "Cash Flow, Consumption Risk, and the Cross-section of Stock Returns," *The Journal of Finance*, 64, 923–956.

Da, Z. and Schaumburg, E. (2011), "Relative valuation and analyst target price forecasts," *Journal of Financial Markets*, 14, 161–192.

Da, Z. and Warachka, M. (2009a), "Long-term earnings growth forecasts, limited attention, and return predictability," in *AFA 2010 Atlanta Meetings Paper*.

Da, Z. and Warachka, M. C. (2009b), "Cashflow risk, systematic earnings revisions, and the cross-section of stock returns," *Journal of Financial Economics*, 94, 448–468.

Da, Z., Liu, Q., and Schaumburg, E. (2011), "Decomposing short-term return reversal," Tech. rep., Staff Report, Federal Reserve Bank of New York.

Daniel, K. and Titman, S. (1997), "Evidence on the characteristics of cross sectional variation in stock returns," *The Journal of Finance*, 52, 1–33.

Daniel, K. and Titman, S. (2006), "Market reactions to tangible and intangible information," *The Journal of Finance*, 61, 1605–1643.

Daniel, K. and Titman, S. (2012), "Testing factor-model explanations of market anomalies," *Critical Finance Review*, 1, 103–139.

Datar, V. T., Y Naik, N., and Radcliffe, R. (1998), "Liquidity and stock returns: An alternative test," *Journal of Financial Markets*, 1, 203–219.

Dichev, I. D. (1998), "Is the risk of bankruptcy a systematic risk?" *the Journal of Finance*, 53, 1131–1147.

Dichev, I. D. and Piotroski, J. D. (2001), "The Long-Run Stock Returns Following Bond Ratings Changes," *The Journal of Finance*, 56, 173–203.

Diether, K. B., Malloy, C. J., and Scherbina, A. (2002), "Differences of opinion and the cross section of stock returns," *The Journal of Finance*, 57, 2113–2141.

Dittmar, R. F. (2002), "Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns," *The Journal of Finance*, 57, 369–403.

Donangelo, A. (2014), "Labor mobility: Implications for asset pricing," *The Journal of Finance*, 69, 1321–1346.

Doran, J., Fodor, A., and Peterson, D. (2007), "Insiders versus Outsiders with Employee Stock Options: Who Knows Best About Future Firm Risk and Implications for Stock Returns," *Available at SSRN 963405*.

Douglas, G. W. (1967), "Risk in the equity markets: An empirical appraisal of market efficiency," Ph.D. thesis, Yale University.

Doyle, J. T., Lundholm, R. J., and Soliman, M. T. (2003), "The predictive value of expenses excluded from pro forma earnings," *Review of Accounting Studies*, 8, 145–174.

Drake, M. S., Rees, L., and Swanson, E. P. (2011), "Should investors follow the prophets or the bears? Evidence on the use of public information by analysts and short sellers," *The Accounting Review*, 86, 101–130.

Drechsler, I. and Yaron, A. (2011), "What's vol got to do with it," *Review of Financial Studies*, 24, 1–45.

Driessen, J. and Maenhout, P. (2007), "An empirical portfolio perspective on option pricing anomalies," *Review of Finance*, 11, 561–603.

Dudoit, S. and Van Der Laan, M. J. (2007), *Multiple testing procedures with applications to genomics*, Springer.

Easley, D., Hvidkjaer, S., and Ohara, M. (2002), "Is information risk a determinant of asset returns?" *The journal of finance*, 57, 2185–2221.

Easley, D., Hvidkjaer, S., and OHara, M. (2010), "Factoring information into returns," .

Eberhart, A. C., Maxwell, W. F., and Siddique, A. R. (2004), "An Examination of Long-Term Abnormal Stock Returns and Operating Performance Following R&D Increases," *The Journal of Finance*, 59, 623–650.

Edmans, A. (2011), "Does the stock market fully value intangibles? Employee satisfaction and equity prices," *Journal of Financial Economics*, 101, 621–640.

Efron, B. (1979), "Bootstrap methods: another look at the jackknife," *The annals of Statistics*, pp. 1–26.

Efron, B. (2004), "Large-scale simultaneous hypothesis testing," *Journal of the American Statistical Association*, 99.

Efron, B. et al. (2008), "Microarrays, empirical Bayes and the two-groups model," *Statistical Science*, 23, 1–22.

Efron, B. and Tibshirani, R. (2002), "Empirical Bayes methods and false discovery rates for microarrays," *Genetic epidemiology*, 23, 70–86.

Efron, B., Tibshirani, R., Storey, J. D., and Tusher, V. (2001), "Empirical Bayes analysis of a microarray experiment," *Journal of the American statistical association*, 96, 1151–1160.

Eiling, E. (2013), "Industry-Specific Human Capital, Idiosyncratic Risk, and the Cross-Section of Expected Stock Returns," *The Journal of Finance*, 68, 43–84.

Eisfeldt, A. L. and Papanikolaou, D. (2013), "Organization Capital and the Cross-Section of Expected Returns," *The Journal of Finance*, 68, 1365–1406.

Elton, E. J., Gruber, M. J., and Blake, C. R. (????).

Elton, E. J., Gruber, M. J., Das, S., and Hlavka, M. (1993), "Efficiency with costly information: A reinterpretation of evidence from managed portfolios," *Review of Financial studies*, 6, 1–22.

Erb, C. B., Harvey, C. R., and Viskanta, T. E. (1996), "Expected returns and volatility in 135 countries," *The Journal of Portfolio Management*, 22, 46–58.

Fairfield, P. M., Whisenant, J. S., and Yohn, T. L. (2003), "Accrued earnings and growth: Implications for future profitability and market mispricing," *The Accounting Review*, 78, 353–371.

Fama, E. F. (1991), "Efficient capital markets: II," *The journal of finance*, 46, 1575–1617.

Fama, E. F. and French, K. R. (1992), "The cross-section of expected stock returns," *the Journal of Finance*, 47, 427–465.

Fama, E. F. and French, K. R. (1993), "Common risk factors in the returns on stocks and bonds," *Journal of financial economics*, 33, 3–56.

Fama, E. F. and French, K. R. (2006), "Profitability, investment and average returns," *Journal of Financial Economics*, 82, 491–518.

Fama, E. F. and French, K. R. (2010), "Luck versus skill in the cross-section of mutual fund returns," *The Journal of Finance*, 65, 1915–1947.

Fama, E. F. and MacBeth, J. D. (1973), "Risk, return, and equilibrium: Empirical tests," *The Journal of Political Economy*, pp. 607–636.

Fang, L. and Peress, J. (2009), "Media Coverage and the Cross-section of Stock Returns," *The Journal of Finance*, 64, 2023–2052.

Farcomeni, A. (2007), "A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion," *Statistical Methods in Medical Research*.

Ferson, W., Nallareddy, S., and Xie, B. (2013), "The out-of-sample performance of long run risk models," *Journal of Financial Economics*, 107, 537–556.

Ferson, W. E. and Harvey, C. R. (1991), "The variation of economic risk premiums," *Journal of Political Economy*, pp. 385–415.

Ferson, W. E. and Harvey, C. R. (1993), "The risk and predictability of international equity returns," *Review of financial Studies*, 6, 527–566.

Ferson, W. E. and Harvey, C. R. (1994), "Sources of risk and expected returns in global equity markets," *Journal of Banking & Finance*, 18, 775–803.

Ferson, W. E. and Harvey, C. R. (1999), "Conditioning variables and the cross section of stock returns," *The Journal of Finance*, 54, 1325–1360.

Ferson, W. E. and Siegel, A. F. (2001), "The efficient use of conditioning information in portfolios," *The Journal of Finance*, 56, 967–982.

Ferson, W. E. and Siegel, A. F. (2003), "Stochastic discount factor bounds with conditioning information," *Review of Financial Studies*, 16, 567–595.

Figlewski, S. (1981), "The informational effects of restrictions on short sales: some empirical evidence," *Journal of Financial and Quantitative Analysis*, 16, 463–476.

Fogler, H. R., JOHN, R., and Tipton, J. (1981), "Three factors, interest rate differentials and stock groups," *The Journal of Finance*, 36, 323–335.

Foster, F. D., Smith, T., and Whaley, R. E. (1997), "Assessing Goodness-of-Fit of Asset Pricing Models: The Distribution of the Maximal R2," *The Journal of Finance*, 52, 591–607.

Frank, M. Z. and Goyal, V. K. (2009), "Capital structure decisions: which factors are reliably important?" *Financial Management*, 38, 1–37.

Frankel, R. and Lee, C. (1998), "Accounting valuation, market expectation, and cross-sectional stock returns," *Journal of Accounting and economics*, 25, 283–319.

Franzoni, F. and Marin, J. M. (2006), "Pension plan funding and stock market efficiency," *the Journal of Finance*, 61, 921–956.

Frazzini, A. and Pedersen, L. H. (2014), "Betting against beta," *Journal of Financial Economics*, 111, 1–25.

Friewald, N., Wagner, C., and Zechner, J. (2014), "The Cross-Section of Credit Risk Premia and Equity Returns," *The Journal of Finance*.

Fu, F. (2009), "Idiosyncratic risk and the cross-section of expected stock returns," *Journal of Financial Economics*, 91, 24–37.

Fung, W. and Hsieh, D. A. (1997), "Empirical characteristics of dynamic trading strategies: The case of hedge funds," *Review of financial studies*, 10, 275–302.

Fung, W. and Hsieh, D. A. (2001), "The risk in hedge fund strategies: Theory and evidence from trend followers," *Review of Financial studies*, 14, 313–341.

Gabaix, X. (2011), "Disasterization: A Simple Way to Fix the Asset Pricing Properties of Macroeconomic Models," *The American Economic Review*, 101, 406–409.

Gallant, A. R., Hansen, L. P., and Tauchen, G. (1990), "Using conditional moments of asset payoffs to infer the volatility of intertemporal marginal rates of substitution," *Journal of Econometrics*, 45, 141–179.

Garcia, D. and Norli, Ø. (2012), "Geographic dispersion and stock returns," *Journal of Financial Economics*, 106, 547–565.

Garlappi, L. and Yan, H. (2011), "Financial Distress and the Cross-section of Equity Returns," *The Journal of Finance*, 66, 789–822.

Garlappi, L., Shu, T., and Yan, H. (2008), "Default risk, shareholder advantage, and stock returns," *Review of Financial Studies*, 21, 2743–2778.

Gârleanu, N., Kogan, L., and Panageas, S. (2012), "Displacement risk and asset returns," *Journal of Financial Economics*, 105, 491–510.

George, T. J. and HWANG, C.-Y. (2004), "The 52-week high and momentum investing," *The Journal of Finance*, 59, 2145–2176.

George, T. J. and Hwang, C.-Y. (2010), "A resolution of the distress risk and leverage puzzles in the cross section of stock returns," *Journal of Financial Economics*, 96, 56–79.

Gettleman, E. and Marks, J. M. (2006), "Acceleration strategies," Tech. rep., Working paper, University of Illinois at Urbana-Champaign.

Gokcen, U. (2009), "Information revelation and expected stock returns," in *FMA 2009 Reno Meetings Paper.*

Gomes, J. F., Yaron, A., and Zhang, L. (2006), "Asset pricing implications of firms financing constraints," *Review of Financial Studies*, 19, 1321–1356.

Gómez, J.-P., Priestley, R., and Zapatero, F. (2012), "Labor income, relative wealth concerns, and the cross-section of stock returns," Tech. rep., Working Paper, Instituto de Empresa Business School.

Gompers, P. A. and Metrick, A. (1998), "Institutional investors and equity prices," Tech. rep., National bureau of economic research.

Gompers, P. A., Ishii, J. L., and Metrick, A. (2001), "Corporate governance and equity prices," Tech. rep., National bureau of economic research.

Gourieroux, C., Holly, A., and Monfort, A. (1982), "Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters," *Econometrica: journal of the Econometric Society*, pp. 63–80.

Gourio, F. (2007), "Labor leverage, firms heterogeneous sensitivities to the business cycle, and the cross-section of expected returns," *Unpublished working paper, Boston University.*

Gourio, F. (2008), "Time-series predictability in the disaster model," *Finance Research Letters*, 5, 191–203.

Gow, I. and Taylor, D. (2009), "Earnings volatility and the cross-section of returns," Tech. rep., Working Paper, Northwestern University.

Green, J., Hand, J., and Zhang, F. (2013a), "The remarkable multidimensionality in the cross section of expected US stock returns," Tech. rep., Working Paper, Pennsylvania State University.

Green, J., Hand, J. R., and Zhang, X. F. (2013b), "The supraview of return predictive signals," *Review of Accounting Studies*, 18, 692–730.

Greene, W. H. (2003), *Econometric analysis*, Pearson Education India.

Griffin, J. M. and Lemmon, M. L. (2002), "Book–to–market equity, distress risk, and stock returns," *The Journal of Finance*, 57, 2317–2336.

Gu, F. (2005), "Innovation, future earnings, and market efficiency," *Journal of Accounting, Auditing & Finance*, 20, 385–418.

Gu, F. and Lev, B. (2011), "Overpriced shares, ill-advised acquisitions, and goodwill impairment," *The Accounting Review*, 86, 1995–2022.

Gu, L., Wang, Z., and Ye, J. (2008), "Information in order backlog: change versus level," Tech. rep., Working Paper, Fordham University.

Guo, H. and Savickas, R. (2008), "Average idiosyncratic volatility in G7 countries," *Review of Financial Studies*, 21, 1259–1296.

Gupta, M. C. and Ofer, A. R. (1975), "Investors' expectations of earnings growth, their accuracy and effects of the structure of realized rates of return," *The Journal of Finance*, 30, 509–523.

Hafzalla, N., Lundholm, R., and Matthew Van Winkle, E. (2011), "Percent accruals," *The Accounting Review*, 86, 209–236.

Hahn, J. and Lee, H. (2009), "Financial Constraints, Debt Capacity, and the Cross-section of Stock Returns," *The Journal of Finance*, 64, 891–921.

Hameed, A., Huang, J., and Mian, G. M. (2010), "Industries and stock return reversals," *SSRN eLibrary*.

Han, B. and Zhou, Y. (2011), "Term structure of credit default swap spreads and cross-section of stock returns," *McCombs Research Paper Series No. FIN-01-11, University of Texas at Austin*.

Han, Y. and Zhou, G. (2013), "Trend factor: A new determinant of cross-section stock returns," *Unpublished working paper, University of Colorado Denver and Washington University in St. Louis*.

Hansen, L. and Scheinkman, J. (2013), "Stochastic Compounding and Uncertain Valuation," *Available at SSRN 2256246*.

Hansen, L. P. and Jagannathan, R. (1991), "Implications of Security Market Data for Models of Dynamic Economies," *The Journal of Political Economy*, 99, 225–262.

Hansen, L. P. and Jagannathan, R. (1997), "Assessing specification errors in stochastic discount factor models," *The Journal of Finance*, 52, 557–590.

Hansen, L. P. and Singleton, K. J. (1983), "Stochastic consumption, risk aversion, and the temporal behavior of asset returns," *The Journal of Political Economy*, pp. 249–265.

Hansen, L. P., Borovička, J., Hendricks, M., and Scheinkman, J. A. (2009), "Risk Price Dynamics," *Becker Friedman Institute for Research in Economics Working Paper*.

Harvey, C. R. and Liu, Y. (2013a), "Backtesting," *Available at SSRN*.

Harvey, C. R. and Liu, Y. (2013b), "Multiple Testing in Economics," *Available at SSRN 2358214*.

Harvey, C. R. and Siddique, A. (2000), "Conditional skewness in asset pricing tests," *The Journal of Finance*, 55, 1263–1295.

Hawkins, E. H., Chamberlin, S. C., and Daniel, W. E. (1984), "Earnings expectations and security prices," *Financial Analysts Journal*, pp. 24–74.

Head, A., Smith, G., and Wilson, J. (2009), "Would a stock by any other ticker smell as sweet?" *The Quarterly Review of Economics and Finance*, 49, 551–561.

Heaton, J. and Lucas, D. (2000), "Portfolio choice and asset prices: The importance of entrepreneurial risk," *The journal of finance*, 55, 1163–1198.

Heckerman, D. G. (1972), "Portfolio selection and the structure of capital asset prices when relative prices of consumption goods may change," *The Journal of Finance*, 27, 47–60.

Heckman, J. J. (1979), "Sample selection bias as a specification error," *Econometrica: Journal of the econometric society*, pp. 153–161.

Hess, D., Kreutzmann, D., and Pucker, O. (2011), "Projected earnings accuracy and the profitability of stock recommendations," Tech. rep., CFR working paper.

Hirshleifer, D. and Jiang, D. (2010), "A financing-based misvaluation factor and the cross-section of expected returns," *Review of Financial Studies*, 23, 3401–3436.

Hirshleifer, D., Hsu, P.-H., and Li, D. (2013), "Innovative efficiency and stock returns," *Journal of Financial Economics*, 107, 632–654.

Hochberg, Y. (1988), "A sharper Bonferroni procedure for multiple tests of significance," *Biometrika*, 75, 800–802.

Hochberg, Y. and Benjamini, Y. (1990), "More powerful procedures for multiple significance testing," *Statistics in medicine*, 9, 811–818.

Hochberg, Y. and Tamhane, A. C. (1987), *Multiple comparison procedures*, John Wiley & Sons, Inc.

Holland, B., Basu, S., and Sun, F. (2010), "Neglect of multiplicity when testing families of related hypotheses," Tech. rep., Working Paper, Temple University.

Holm, S. (1979), "A simple sequentially rejective multiple test procedure," *Scandinavian journal of statistics*, pp. 65–70.

Holthausen, R. W. and Larcker, D. F. (1992), "The prediction of stock returns using financial statement information," *Journal of Accounting and Economics*, 15, 373–411.

Hommel, G. (1988), "A stagewise rejective multiple test procedure based on a modified Bonferroni test," *Biometrika*, 75, 383–386.

Hou, K. and Moskowitz, T. J. (2005), "Market frictions, price delay, and the cross-section of expected returns," *Review of Financial Studies*, 18, 981–1020.

Hou, K. and Robinson, D. T. (2006), "Industry concentration and average stock returns," *The Journal of Finance*, 61, 1927–1956.

Hou, K., Karolyi, G. A., and Kho, B.-C. (2011), "What factors drive global stock returns?" *Review of Financial Studies*, 24, 2527–2574.

Hu, G. X., Pan, J., and Wang, J. (2013), "Noise as information for illiquidity," *The Journal of Finance*, 68, 2341–2382.

Huang, A. G. (2009), "The cross section of cashflow volatility and expected stock returns," *Journal of Empirical Finance*, 16, 409–429.

Huang, W., Liu, Q., Ghon Rhee, S., and Wu, F. (2012), "Extreme downside risk and expected stock returns," *Journal of Banking & Finance*, 36, 1492–1502.

Hvidkjaer, S. (2008), "Small trades and the cross-section of stock returns," *Review of Financial Studies*.

Imrohoroglu, A. and Tüzel, S. (2014), "Firm-Level Productivity, Risk, and Return," *Management Science*.

Ioannidis, J. P. (2005), "Why most published research findings are false," *PLoS medicine*, 2, e124.

Jackwerth, J. C. (2000), "Recovering risk aversion from option prices and realized returns," *Review of Financial Studies*, 13, 433–451.

Jacobs, K. and Wang, K. Q. (2004), "Idiosyncratic consumption risk and the cross section of asset returns," *The Journal of Finance*, 59, 2211–2252.

Jagannathan, R. and Marakani, S. (2011), *Long Run Risks & Price/Dividend Ratio Factors*, National Bureau of Economic Research.

Jagannathan, R. and Wang, Y. (2007), "Lazy Investors, Discretionary Consumption, and the Cross-Section of Stock Returns," *The Journal of Finance*, 62, 1623–1661.

Jagannathan, R. and Wang, Z. (1996), "The conditional CAPM and the cross-section of expected returns," *The Journal of Finance*, 51, 3–53.

Jarrow, R. (1980), "Heterogeneous expectations, restrictions on short sales, and equilibrium asset prices," *The Journal of Finance*, 35, 1105–1113.

Jefferys, W. H. and Berger, J. O. (1992), "Ockham's razor and Bayesian analysis," *American Scientist*, pp. 64–72.

Jegadeesh, N. (1990), "Evidence of predictable behavior of security returns," *The Journal of Finance*, 45, 881–898.

Jegadeesh, N. and Titman, S. (1993), "Returns to buying winners and selling losers: Implications for stock market efficiency," *The Journal of Finance*, 48, 65–91.

Jegadeesh, N., Kim, J., Krische, S. D., and Lee, C. (2004), "Analyzing the analysts: When do recommendations add value?" *The journal of finance*, 59, 1083–1124.

Jensen, M. C., Black, F., and Scholes, M. S. (1972), "The capital asset pricing model: Some empirical tests," .

Jiang, G., Lee, C. M., and Zhang, Y. (2005), "Information uncertainty and expected returns," *Review of Accounting Studies*, 10, 185–221.

Jiang, H. and Sun, Z. (2011), "Dispersion in beliefs among active mutual funds and the cross-section of stock returns," Tech. rep., Working Paper, Erasmus University.

Johnson, T. L. and So, E. C. (2012), "The option to stock volume ratio and future returns," *Journal of Financial Economics*, 106, 262–286.

Jones, C. M. and Lamont, O. A. (2002), "Short-sale constraints and stock returns," *Journal of Financial Economics*, 66, 207–239.

Kapadia, N. (2011), "Tracking down distress risk," *Journal of Financial Economics*, 102, 167–182.

Kaplan, S. N. and Zingales, L. (1997), "Do investment-cash flow sensitivities provide useful measures of financing constraints?" *The Quarterly Journal of Economics*, pp. 169–215.

Kelly, B. and Pruitt, S. (2011), "The three-pass regression filter: A new approach to forecasting using many predictors," *Fama & Miller Working Paper*.

Kim, C. F., Pantzalis, C., and Chul Park, J. (2012), "Political geography and stock returns: The value and risk implications of proximity to political power," *Journal of Financial Economics*, 106, 196–228.

Koijen, R. S., Moskowitz, T. J., Pedersen, L. H., and Vrugt, E. B. (2013), "Carry," Tech. rep., National Bureau of Economic Research.

Korajczyk, R. A. and Sadka, R. (2008), "Pricing the commonality across alternative measures of liquidity," *Journal of Financial Economics*, 87, 45–72.

Korniotis, G. M. (2008), "Habit formation, incomplete markets, and the significance of regional risk for expected returns," *Review of Financial Studies*, 21, 2139–2172.

Korniotis, G. M. and Kumar, A. (2009), "Long Georgia, short Colorado? The geography of return predictability," *The Geography of Return Predictability (January 16, 2009)*.

Kosowski, R., Timmermann, A., Wermers, R., and White, H. (2006), "Can mutual fund stars really pick stocks? New evidence from a bootstrap analysis," *The Journal of finance*, 61, 2551–2595.

Kosowski, R., Naik, N. Y., and Teo, M. (2007), "Do hedge funds deliver alpha? A Bayesian and bootstrap analysis," *Journal of Financial Economics*, 84, 229–264.

Kraus, A. and Litzenberger, R. H. (1976), "Skewness preference and the valuation of risk assets," *The Journal of Finance*, 31, 1085–1100.

Kumar, A. and Lee, C. (2006), "Retail investor sentiment and return comovements," *The Journal of Finance*, 61, 2451–2486.

Kumar, P., Sorescu, S. M., Boehme, R. D., and Danielsen, B. R. (2008), "Estimation risk, information, and the conditional CAPM: theory and evidence," *Review of Financial Studies*, 21, 1037–1075.

Kyle, A. S. (1985), "Continuous auctions and insider trading," *Econometrica: Journal of the Econometric Society*, pp. 1315–1335.

Lamont, O., Polk, C., and Saa-Requejo, J. (2001), "Financial constraints and stock returns," *Review of financial studies*, 14, 529–554.

Landsman, W. R., Miller, B. L., Peasnell, K., and Yeh, S. (2011), "Do investors understand really dirty surplus?" *The Accounting Review*, 86, 237–258.

Larcker, D. F., So, E. C., and Wang, C. C. (2013), "Boardroom centrality and firm performance," *Journal of Accounting and Economics*, 55, 225–250.

Leamer, E. E. and Leamer, E. E. (1978), *Specification searches: Ad hoc inference with nonexperimental data*, Wiley New York.

Lee, C. and Swaminathan, B. (2000), "Price momentum and trading volume," *The Journal of Finance*, 55, 2017–2069.

Lehavy, R. and Sloan, R. G. (2008), "Investor recognition and stock returns," *Review of Accounting Studies*, 13, 327–361.

Lehmann, E. L. and Romano, J. P. (2012), *Generalizations of the familywise error rate*, Springer.

Lettau, M. and Ludvigson, S. (2001), "Resurrecting the (C) CAPM: A cross-sectional test when risk premia are time-varying," *Journal of Political Economy*, 109, 1238–1287.

Lev, B. and Sougiannis, T. (1996), "The capitalization, amortization, and value-relevance of R&D," *Journal of accounting and economics*, 21, 107–138.

Lev, B., Nissim, D., and Thomas, J. (2002), "On the informational usefulness of R&D capitalization and amortization," *Pages. stern. nyu. edu.*

Lev, B., Sarath, B., and Sougiannis, T. (2005), "R&D Reporting Biases and Their Consequences*," *Contemporary Accounting Research*, 22, 977–1026.

Lewellen, J., Nagel, S., and Shanken, J. (2010), "A skeptical appraisal of asset pricing tests," *Journal of Financial Economics*, 96, 175–194.

Li, D. (2011a), "Financial constraints, R&D investment, and stock returns," *Review of Financial Studies*, p. hhr043.

Li, K. K. (2011b), "How well do investors understand loss persistence?" *Review of Accounting Studies*, 16, 630–667.

Li, Q., Vassalou, M., and Xing, Y. (2006), "Sector Investment Growth Rates and the Cross Section of Equity Returns," *The Journal of Business*, 79, 1637–1665.

Li, S. Z. (2012), "Continuous beta, discontinuous beta, and the cross-section of expected stock returns," Tech. rep., Working Paper, Duke University.

Li, X. (2010), "Real earnings management and subsequent stock returns," *Available at SSRN*, 1679832.

Liang, Y., Kelemen, A., et al. (2008), "Statistical advances and challenges for analyzing correlated high dimensional SNP data in genomic study for complex diseases," *Statistics Surveys*, 2, 43–60.

Lintner, J. (1965), "Security Prices, Risk, and Maximal Gains from Diversification," *The Journal of Finance*, 20, 587–615.

Lioui, A. and Maio, P. (2012), "Interest rate risk and the cross-section of stock returns," *Journal of Financial and Quantitative Analysis*, pp. 1–48.

Litzenberger, R. H. and Ramaswamy, K. (1979), "The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence," *Journal of financial economics*, 7, 163–195.

Liu, W. (2006), "A liquidity-augmented capital asset pricing model," *Journal of financial Economics*, 82, 631–671.

Liu, Y. (2012), "Index option returns and generalized entropy bounds," *Working Paper, Duke University.*

Liu, Y. (2013), "Diagnosing Dynamic Asset Pricing Models Using Generalized Entropy Bounds," *Working Paper, Duke University.*

Livdan, D., Sapriza, H., and Zhang, L. (2009), "Financially constrained stock returns," *The Journal of Finance*, 64, 1827–1862.

Lo, A. W. and MacKinlay, A. C. (1990), "Data-snooping biases in tests of financial asset pricing models," *Review of financial studies*, 3, 431–467.

Lo, A. W. and Wang, J. (2006), "Trading volume: Implications of an intertemporal capital asset pricing model," *The Journal of Finance*, 61, 2805–2840.

Longstaff, F. A. and Piazzesi, M. (2004), "Corporate earnings and the equity premium," *Journal of Financial Economics*, 74, 401–421.

Loughran, T. and Ritter, J. R. (1995), "The new issues puzzle," *The Journal of Finance*, 50, 23–51.

Loughran, T. and Vijh, A. M. (1997), "Do long-term shareholders benefit from corporate acquisitions?" *The Journal of Finance*, 52, 1765–1790.

Lucas Jr, R. E. (1978), "Asset prices in an exchange economy," *Econometrica: Journal of the Econometric Society*, pp. 1429–1445.

Lustig, H. N. and Van Nieuwerburgh, S. G. (2005), "Housing collateral, consumption insurance, and risk premia: An empirical perspective," *The Journal of Finance*, 60, 1167–1219.

Lynch, A. W. and Vital-Ahuja, T. (1998), "Can subsample evidence alleviate the data-snooping problem? A comparison to the maximal R2 cutoff test," Tech. rep., Discussion paper, Stern Business School, New York University.

Malloy, C. J., Moskowitz, T. J., and VISSING-JØRGENSEN, A. (2009), "Long-Run Stockholder Consumption Risk and Asset Returns," *The Journal of Finance*, 64, 2427–2479.

Martin, I. (2007), "Disasters and asset pricing: evidence from option markets," *manuscript, February.*

Martin, I. W. (2008), "Disasters and the welfare cost of uncertainty," *The American Economic Review*, pp. 74–78.

Martin, I. W. (2013), "Consumption-based asset pricing with higher cumulants," *The Review of Economic Studies*, 80, 745–773.

Mayshar, J. (1981), "Transaction costs and the pricing of assets," *The journal of Finance*, 36, 583–597.

McConnell, J. J. and Sanger, G. C. (1984), "A trading strategy for new listings on the NYSE," *Financial Analysts Journal*, pp. 34–38.

McLean, R. D. and Pontiff, J. (2012), "Does academic research destroy stock return predictability," *SSRN eLibrary.*

Meinshausen, N. (2008), "Hierarchical testing of variable importance," *Biometrika*, 95, 265–278.

Meng, C. Y. and Dempster, A. P. (1987), "A Bayesian approach to the multiplicity problem for significance testing with binomial data," *Biometrics*, pp. 301–311.

Menzly, L. and Ozbas, O. (2010), "Market Segmentation and Cross-predictability of Returns," *The Journal of Finance*, 65, 1555–1580.

Merton, R. C. (1973), "An intertemporal capital asset pricing model," *Econometrica: Journal of the Econometric Society*, pp. 867–887.

Michaely, R., Thaler, R. H., and Womack, K. L. (1995), "Price reactions to dividend initiations and omissions: overreaction or drift?" *The Journal of Finance*, 50, 573–608.

Mohanram, P. S. (2005), "Separating Winners from Losers among LowBook-to-Market Stocks using Financial Statement Analysis," *Review of Accounting Studies*, 10, 133–170.

Moskowitz, T. J. and Grinblatt, M. (1999), "Do industries explain momentum?" *The Journal of Finance*, 54, 1249–1290.

Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H. (2012), "Time series momentum," *Journal of Financial Economics*, 104, 228–250.

Mossin, J. (1966), "Equilibrium in a capital asset market," *Econometrica: Journal of the econometric society*, pp. 768–783.

Nagel, S. (2005), "Short sales, institutional investors and the cross-section of stock returns," *Journal of Financial Economics*, 78, 277–309.

Naik, V. and Lee, M. (1990), "General equilibrium pricing of options on the market portfolio with discontinuous returns," *Review of Financial Studies*, 3, 493–521.

Narayanamoorthy, G. (2006), "Conservatism and Cross-Sectional Variation in the Post–Earnings Announcement Drift," *Journal of Accounting Research*, 44, 763–789.

Nguyen, G. X. and Swanson, P. E. (2009), "Firm characteristics, relative efficiency, and equity returns," *Journal of Financial and Quantitative Analysis*, 44, 213–236.

Novy-Marx, R. (2013), "The other side of value: The gross profitability premium," *Journal of Financial Economics*, 108, 1–28.

Nyberg, P. and Pöyry, S. (2014), "Firm Expansion and Stock Price Momentum*," *Review of Finance*, 18, 1465–1505.

Ofek, E., Richardson, M., and Whitelaw, R. F. (2004), "Limited arbitrage and short sales restrictions: Evidence from the options markets," *Journal of Financial Economics*, 74, 305–342.

Oldfield, G. S. and Rogalski, R. J. (1981), "Treasury bill factors and common stock returns," *The Journal of Finance*, 36, 337–350.

Ortiz-Molina, H. and Phillips, G. M. (2010), "Asset liquidity and the cost of capital," Tech. rep., National Bureau of Economic Research.

Ou, J. A. and Penman, S. H. (1989), "Financial statement analysis and the prediction of stock returns," *Journal of accounting and economics*, 11, 295–329.

Palazzo, B. (2012), "Cash holdings, risk, and expected returns," *Journal of Financial Economics*, 104, 162–185.

Pan, J. (2002), "The jump-risk premia implicit in options: Evidence from an integrated time-series study," *Journal of financial economics*, 63, 3–50.

Papanastasopoulos, G., Thomakos, D., and Wang, T. (2010), "The implications of retained and distributed earnings for future profitability and stock returns," *Review of Accounting and Finance*, 9, 395–423.

Parker, J. A. and Julliard, C. (2005), "Consumption risk and the cross section of expected returns," *Journal of Political Economy*, 113, 185–222.

Pastor, L. and Stambaugh, R. F. (2001), "Liquidity risk and expected stock returns," Tech. rep., National Bureau of Economic Research.

Patatoukas, P. N. (2011), "Customer-Base Concentration: Implications for Firm Performance and Capital Markets: 2011 American Accounting Association Competitive Manuscript Award Winner," *The Accounting Review*, 87, 363–392.

Patton, A. J. and Timmermann, A. (2010), "Monotonicity in asset returns: New tests with applications to the term structure, the CAPM, and portfolio sorts," *Journal of Financial Economics*, 98, 605–625.

Penman, S. and Zhang, X.-J. (2002), "Modeling Sustainable earnings and P/E ratios with financial statement analysis," *SSRN eLibrary(June 1). http://papers. ssrn. com/sol3/papers. cfm.*

Pesaran, M. H. and Timmermann, A. (2007), "Selection of estimation window in the presence of breaks," *Journal of Econometrics*, 137, 134–161.

Petkova, R., Akbas, F., and Armstrong, W. J. (2011), "The Volatility of Liquidity and Expected Stock Returns," *Available at SSRN 1786991.*

Piotroski, J. D. (2000), "Value investing: The use of historical financial statement information to separate winners from losers," *Journal of Accounting Research*, pp. 1–41.

Pontiff, J. and Woodgate, A. (2008), "Share issuance and cross-sectional returns," *The Journal of Finance*, 63, 921–945.

PORTA, R. (1996), "Expectations and the cross-section of stock returns," *The Journal of Finance*, 51, 1715–1742.

Prakash, R. and Sinha, N. (2013), "Deferred Revenues and the Matching of Revenues and Expenses*," *Contemporary Accounting Research*, 30, 517–548.

Price, S. M., Doran, J. S., Peterson, D. R., and Bliss, B. A. (2012), "Earnings conference calls and stock returns: The incremental informativeness of textual tone," *Journal of Banking & Finance*, 36, 992–1011.

Rajgopal, S., Shevlin, T., and Venkatachalam, M. (2003), "Does the stock market fully appreciate the implications of leading indicators for future earnings? Evidence from order backlog," *Review of Accounting Studies*, 8, 461–492.

Rietz, T. A. (1988), "The equity risk premium a solution," *Journal of monetary Economics*, 22, 117–131.

Roll, R. (1988), "R^2," *Journal of Finance*, 43, 541–566.

Romano, J. P., Shaikh, A. M., and Wolf, M. (2008a), "Control of the false discovery rate under dependence using the bootstrap and subsampling," *Test*, 17, 417–442.

Romano, J. P., Shaikh, A. M., and Wolf, M. (2008b), "Formalized data snooping based on generalized error rates," *Econometric Theory*, 24, 404–447.

Rosenthal, R. (1979), "The file drawer problem and tolerance for null results." *Psychological bulletin*, 86, 638.

Ross, S. A. (1987), "Regression to the Max," *Unpublished Paper, Yale University.*

Rubinstein, M. (1974), "An aggregation theorem for securities markets," *Journal of Financial Economics*, 1, 225–244.

Rubinstein, M. E. (1973), "The fundamental theorem of parameter-preference security valuation," *Journal of Financial and Quantitative Analysis*, 8, 61–69.

Sadka, R. (2006), "Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk," *Journal of Financial Economics*, 80, 309–349.

Santos, T. and Veronesi, P. (2010), "Habit formation, the cross section of stock returns and the cash-flow risk puzzle," *Journal of Financial Economics*, 98, 385–413.

Sarkar, S. K. (2002), "Some results on false discovery rate in stepwise multiple testing procedures," *Annals of statistics*, pp. 239–257.

Sarkar, S. K. and Guo, W. (2009), "On a generalized false discovery rate," *The Annals of Statistics*, pp. 1545–1565.

Saville, D. J. (1990), "Multiple comparison procedures: the practical solution," *The American Statistician*, 44, 174–180.

Savov, A. (2011), "Asset pricing with garbage," *The Journal of Finance*, 66, 177–201.

Scheffe, H. (1999), *The analysis of variance*, vol. 72, John Wiley & Sons.

Schweder, T. and Spjøtvoll, E. (1982), "Plots of p-values to evaluate many tests simultaneously," *Biometrika*, 69, 493–502.

Schwert, G. W. (2003), "Anomalies and market efficiency," *Handbook of the Economics of Finance*, 1, 939–974.

Scott, J. G. (2009), "Bayesian adjustment for multiplicity," Ph.D. thesis, Duke University.

Scott, J. G. and Berger, J. O. (2006), "An exploration of aspects of Bayesian multiple testing," *Journal of Statistical Planning and Inference*, 136, 2144–2162.

Scott, J. G., Berger, J. O., et al. (2010), "Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem," *The Annals of Statistics*, 38, 2587–2619.

Shaffer, J. P. (1995), "Multiple hypothesis testing," *Annual review of psychology*, 46, 561–584.

Shanken, J. (1990), "Intertemporal asset pricing: An empirical investigation," *Journal of Econometrics*, 45, 99–120.

Sharpe, W. F. (1964), "Capital asset prices: A theory of market equilibrium under conditions of risk*," *The journal of finance*, 19, 425–442.

Shu, T. (2006), "Trader composition, price efficiency, and the cross-section of stock returns," Tech. rep., Working Paper, University of Texas at Austin.

Simes, R. J. (1986), "An improved Bonferroni procedure for multiple tests of significance," *Biometrika*, 73, 751–754.

Simutin, M. (2010), "Excess cash and stock returns," *Financial Management*, 39, 1197–1222.

Sloan, R. G. (1996), "Do stock prices fully reflect information in accruals and cash flows about future earnings?" *Accounting Review*, pp. 289–315.

Snow, K. N. (1991), "Diagnosing asset pricing models using the distribution of asset returns," *The Journal of Finance*, 46, 955–983.

So, E. C. (2013), "A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts?" *Journal of Financial Economics*, 108, 615–640.

Soliman, M. T. (2008), "The use of DuPont analysis by market participants," *The Accounting Review*, 83, 823–853.

Spiess, D. K. and Affleck-Graves, J. (1995), "Underperformance in long-run stock returns following seasoned equity offerings," *Journal of Financial Economics*, 38, 243–267.

Spiess, D. K. and Affleck-Graves, J. (1999), "The long-run performance of stock returns following debt offerings," *Journal of Financial Economics*, 54, 45–73.

Storey, J. D. (2003), "The positive false discovery rate: A Bayesian interpretation and the q-value," *Annals of statistics*, pp. 2013–2035.

Stulz, R. (1981), "A model of international asset pricing," *Journal of Financial Economics*, 9, 383–406.

Stulz, R. M. (1986), "Asset pricing and expected inflation," *The Journal of Finance*, 41, 209–223.

Stutzer, M. (1995), "A Bayesian approach to diagnosis of asset pricing models," *Journal of Econometrics*, 68, 367–397.

Subrahmanyam, A. (2010), "The Cross-Section of Expected Stock Returns: What Have We Learnt from the Past Twenty-Five Years of Research?" *European Financial Management*, 16, 27–42.

Sullivan, R., Timmermann, A., and White, H. (1999), "Data-snooping, technical trading rule performance, and the bootstrap," *The journal of Finance*, 54, 1647–1691.

Sullivan, R., Timmermann, A., and White, H. (2001), "Dangers of data mining: The case of calendar effects in stock returns," *Journal of Econometrics*, 105, 249–286.

Sweeney, R. J. and Warga, A. D. (1986), "The Pricing of Interest-Rate Risk: Evidence from the Stock Market," *The Journal of Finance*, 41, 393–410.

Teo, M. and Woo, S.-J. (2004), "Style effects in the cross-section of stock returns," *Journal of Financial Economics*, 74, 367–398.

Thomas, J. and Zhang, F. X. (2011), "Tax expense momentum," *Journal of Accounting Research*, 49, 791–821.

Thornton, A. and Lee, P. (2000), "Publication bias in meta-analysis: its causes and consequences," *Journal of clinical epidemiology*, 53, 207–216.

Titman, S., Wei, K.-C., and Xie, F. (2004), "Capital investments and stock returns," *Journal of Financial and Quantitative Analysis*, 39, 677–700.

Todorov, V. and Bollerslev, T. (2010), "Jumps and betas: A new framework for disentangling and estimating systematic risks," *Journal of Econometrics*, 157, 220–235.

Troendle, J. F. (2000), "Stepwise normal theory multiple test procedures controlling the false discovery rate," *Journal of Statistical Planning and Inference*, 84, 139–158.

Valta, P. (2013), "Strategic default, debt structure, and stock returns," *Debt Structure, and Stock Returns (March 1, 2013).*

Van Binsbergen, J. H. (2007), "Good-specific habit formation and the cross-section of expected returns," in *AFA 2009 San Francisco Meetings Paper*.

Vanden, J. M. (2004), "Options Trading and the CAPM," *Review of Financial Studies*, 17, 207–238.

Vanden, J. M. (2006), "Option coskewness and capital asset pricing," *Review of Financial Studies*, 19, 1279–1320.

Vassalou, M. (????).

Vassalou, M. and Xing, Y. (2004), "Default risk in equity returns," *The Journal of Finance*, 59, 831–868.

Viale, A. M., Garcia-Feijoo, L., and Giannetti, A. (2012), "Safety first, robust dynamic asset pricing, and the cross-section of expected stock returns," Tech. rep., Working Paper, Florida Atlantic University.

Wachter, J. A. (2013), "Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?" *The Journal of Finance*, 68, 987–1035.

Wagenmakers, E.-J. and Grünwald, P. (2006), "A Bayesian Perspective on Hypothesis Testing A Comment on Killeen (2005)," *Psychological Science*, 17, 641–642.

Wahlen, J. M. and Wieland, M. M. (2011), "Can financial statement analysis beat consensus analysts recommendations?" *Review of Accounting Studies*, 16, 89–115.

Wang, Y. (2012), "Debt covenants and cross-sectional equity returns," Tech. rep., Working Paper, Concordia University.

Watkins, B. (2003), "Riding the wave of sentiment: An analysis of return consistency as a predictor of future returns," *The Journal of Behavioral Finance*, 4, 191–200.

Welch, I. and Goyal, A. (2008), "A comprehensive look at the empirical performance of equity premium prediction," *Review of Financial Studies*, 21, 1455–1508.

Westfall, P. H. (1993), *Resampling-based multiple testing: Examples and methods for p-value adjustment*, vol. 279, John Wiley & Sons.

White, H. (2000), "A reality check for data snooping," *Econometrica*, 68, 1097–1126.

Whited, T. M. and Wu, G. (2006), "Financial constraints risk," *Review of Financial Studies*, 19, 531–559.

Whittemore, A. S. (2007), "A Bayesian false discovery rate for multiple testing," *Journal of Applied Statistics*, 34, 1–9.

Wolak, F. A. (1987), "An exact test for multiple inequality and equality constraints in the linear regression model," *Journal of the American Statistical Association*, 82, 782–793.

Womack, K. L. (1996), "Do brokerage analysts' recommendations have investment value?" *The Journal of Finance*, 51, 137–167.

Xing, Y. (2008), "Interpreting the value effect through the Q-theory: An empirical investigation," *Review of Financial Studies*, 21, 1767–1795.

Xing, Y., Zhang, X., and Zhao, R. (2010), "What does the individual option volatility smirk tell us about future equity returns?" .

Yan, S. (2011), "Jump risk, stock returns, and slope of implied volatility smile," *Journal of Financial Economics*, 99, 216–233.

Yekutieli, D. and Benjamini, Y. (1999), "Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics," *Journal of Statistical Planning and Inference*, 82, 171–196.

Yogo, M. (2006), "A consumption-based explanation of expected stock returns," *The Journal of Finance*, 61, 539–580.

Zehetmayer, S. and Posch, M. (2010), "Post hoc power estimation in large-scale multiple testing problems," *Bioinformatics*, 26, 1050–1056.

Zhang, X. (2006), "Information uncertainty and stock returns," *The Journal of Finance*, 61, 105–137.

Zhao, X. (2012), “Information intensity and the cross-section of stock returns,” *Working Paper*.

Biography

Yan Liu was born in Harbin, Heilongjiang, China on July 31st, 1983. He received his Bachelor's degree in Mathematics from Tsinghua University in July 2006. Afterwards, he came to the statistics program at the University of Minnesota to pursue a PhD degree. He earned his master's degree in statistics after two years and left to pursue a PhD in finance at Duke University in August 2008.

During his study at Tsinghua University, Yan won Academic Excellence Scholarships in three consecutive years and graduated with Distinction. He also received full scholarship at the University of Minnesota and fellowship at Duke University. Yan's current research focuses on nonparametric asset pricing bounds and multiple hypothesis testing in financial economics. His recent work on the cross-section of expected returns, jointly with Campbell R. Harvey and Heqing Zhu, has won the Best Paper Award for the INQUIRE-Europe-UK 2014 Conference and the NASDAQ OMX Award for the best paper in asset pricing at the 2014 Western Finance Association Meetings.